

Dall'Informatica umanistica alle culture digitali

In memoria di Giuseppe Gigliozzi



Collezione di "Testi e Studi Umanistici"

Serie

Letterature, linguaggi e culture

Quaderni Digilab, 2

Il convegno si è svolto con il patrocinio
dell'Associazione per l'Informatica Umanistica e la Cultura Digitale

Quaderni Digilab

Editorial Board

Fabio CIOTTI, Università di Roma Tor Vergata

Gianfranco CRUPI, Sapienza Università di Roma

Domenico FIORMONTE, Università di Roma Tre

Mariella GUERCIO, Sapienza Università di Roma

Giovanni RAGONE, Sapienza Università di Roma

Mario RICCIARDI, Università di Torino

Gino RONCAGLIA, Università della Tuscia

Giovanni SOLIMINE, Sapienza Università di Roma

Dall'Informatica umanistica alle culture digitali

Atti del convegno di studi (Roma, 27-28 ottobre 2011)
in memoria di Giuseppe Gigliozzi

a cura di
Fabio Ciotti e Gianfranco Crupi



SAPIENZA
UNIVERSITÀ EDITRICE

digitLab
ricerca e servizi

2012

Quaderni DigiLab are published by **Casa Editrice Università La Sapienza**

Published online in **2012**

The copyediting is by **Digital Publishing Division of DigiLab**

(Centro interdipartimentale di ricerca e servizi)

La Sapienza Università di Roma

Managing editor: **Gianfranco Crupi**

Copyediting and Layout: **Ivan Macculi**



SAPIENZA
UNIVERSITÀ EDITRICE



This work is licensed under a Creative Commons 3.0 License



DOI: <http://dx.doi.org/10.7357/DigiLab-32>

Indice

<i>Prefazione</i>	9
Fabio Ciotti, <i>A Giuseppe, amico e maestro</i>	19
Lo statuto dell'informatica umanistica	
Giovanni Ragone, <i>Introduzione ai lavori</i>	24
Tito Orlandi, <i>Per una storia dell'informatica umanistica</i>	49
Dino Buzzetti, <i>Che cos'è, oggi, l'informatica umanistica? L'impatto della tecnologia</i>	103
Il testo e il computer	
Maurizio Lana, <i>Da una digital library del latino tardo ad un corpus globale</i>	134
Paolo Mastandrea, <i>1991-2011, un ventennio di evoluzione nella filologia pratica</i>	151
Raul Mordenti, <i>Domande teoriche sul concetto di edizione (nel nome di Giuseppe Gigliozzi)</i>	167
Federica Perazzini, <i>Oltre i confini del testo: le Digital Humanities tra scienza e opportunità</i>	183
Francesca Tomasi, <i>Digital editions between embedded markup and external representation. A case study: Vespasiano da Bisticci's Letters</i>	201
Oltre l'informatica umanistica: culture digitali e reti della conoscenza	
Domenico Fiormonte, <i>Per un critica culturale delle Digital Humanities</i>	220
Fabio Ciotti, <i>Web semantico, linked data e studi letterari: verso una nuova convergenza</i>	243
Gianfranco Crupi, <i>Universo bibliografico e semantic web</i>	277
Emiliano Ilardi, <i>La narrazione in rete dei patrimoni culturali digitalizzati</i>	307
Christoph Bläsi, <i>The book as a medium: interesting phenomena to be researched with the help of digital humanities approaches</i>	319
Stefano Lariccia, <i>Automi e linguaggio nell'ecosistema delle reti digitali</i>	339

Prefazione

Lo scorso 28 ottobre 2011 ricorreva il decennale della scomparsa di Giuseppe Gigliozzi. La sua figura di studioso e di docente ha rappresentato un punto di riferimento fondamentale per lo sviluppo dell'Informatica umanistica in Italia: con il suo lavoro ha contribuito, infatti, sia alla riflessione teorica, sia alla dimensione sperimentale e applicativa di questo campo disciplinare. I suoi interessi scientifici sono stati amplissimi, spaziando dal mondo, a lui più vicino, degli studi critico-letterari con metodologie informatiche, che lo portarono prima a esplorare le potenzialità dell'analisi automatica del racconto mediante tecnologie mutuatae dall'intelligenza artificiale e successivamente all'uso di strumenti di analisi quantitativa del testo nel lavoro critico¹; al problema teorico e pratico della codifica digitale dei testi; a quello degli studi di archivistica e biblioteconomia digitale, dove fu pioniere nella riflessione e nella sperimentazione sulle biblioteche digitali; fino alle prime riflessioni sulle nuove forme delle culture di rete e della testualità digitale. L'antologia dei suoi lavori raccolti da Myriam Trevisan in *Saggi di informatica umanistica* possono fornire una idea di questa vitalità e profondità teorica². Ma probabilmente l'eredità più importante che Gigliozzi ha lasciato consiste nelle diverse generazioni di ricercatori che passando per il

¹ Capitale resta la sua analisi di *Memoriale* di Paolo Volponi pubblicata nella Letteratura Italiana Einaudi: "Memoriale di Paolo Volponi". In *Letteratura Italiana. Le opere*, diretta da Alberto Asor Rosa, IV. Il Novecento, II. La ricerca letteraria, Torino, Einaudi, 1996, pp. 729-769.

²Giuseppe Gigliozzi, *Saggi di informatica umanistica*, Milano, UNICOPLI, 2008.

suo magistero hanno contribuito alla crescita dell'Informatica umanistica in Italia e non solo.

Per ricordare la sua figura, DigiLab - Centro interdipartimentale di ricerca e servizi dell'Università di Roma La Sapienza e la Facoltà di Lettere e Filosofia dell'Università Roma Tor Vergata, con il patrocinio dell'Associazione per l'Informatica Umanistica e la Cultura Digitale, hanno organizzato un convegno di studi che si è tenuto il 27 e 28 ottobre del 2011. Questo volume raccoglie sulla pagina scritta gli interventi più significativi di quel convegno, ma forse senza riuscire a rendere conto fino in fondo del clima di dibattito e di partecipazione che lo ha caratterizzato.

I lavori si sono articolati in tre sessioni tematiche, che questi atti rispecchiano nella struttura. La prima sezione si intitola *Lo statuto dell'informatica umanistica* e i contributi, che la costituiscono, tentano di definire che cosa caratterizza oggi la famiglia di saperi e pratiche nota in Italia come informatica umanistica, e di individuare il suo ruolo nel contesto delle discipline umanistiche in generale. L'informatica umanistica si è costituita come un complesso di teorie, ricerche e applicazioni elaborati a partire dagli anni settanta del Novecento; il dibattito circa il suo statuto e la sua natura ne ha accompagnato lo sviluppo, ed è stato caratterizzato da forti polarizzazioni oppostive, una dialettica che negli ultimi anni ha ripreso vigore. Nel contesto italiano, in particolare, il discorso sui fondamenti è stato particolarmente vivace e stimolante e la comunità di studio italiana, con in testa la "scuola romana" di cui Giuseppe Gigliozzi è stato uno dei leader più apprezzati, è stata pioniera in questo campo, producendo sia raffinati contributi teorici sia soluzioni tecniche all'avanguardia. Anche dal punto di vista didattico sono nate numerose iniziative e, fino alle recenti riforme, l'università italiana era in Europa, insieme al Regno Unito, la sola a

offrire sia corsi di laurea triennale sia magistrale. A un panorama così variegato e vitale tuttavia non ha fatto da riscontro un riconoscimento istituzionale, ovvero una base solida per garantire all'informatica umanistica continuità e sviluppo. Ebbene, gli interventi inclusi in questa sezione cercano di fare il punto su tali questioni, proponendo nuove idee per consolidare e rilanciare questa disciplina.

Il saggio di Giovanni Ragone (ideatore e fondatore del DigiLab), *Introduzione ai lavori*, descrive lo "scenario di fondo" su cui si colloca il dibattito relativo all'informatica umanistica, riflettendo sul ruolo che le istituzioni culturali debbono assumere nella svolta digitale, e sul cambiamento richiesto all'università per far fronte alle nuove richieste nel settore della formazione e della ricerca.

Il contributo di Tito Orlandi, a cui va riconosciuto il ruolo di vero e proprio fondatore dell'informatica umanistica in Italia, è intitolato *Per una storia dell'informatica umanistica* e propone una dettagliata e documentata ricostruzione storica delle attività della cosiddetta "scuola romana" nel decennio '80-'90, nel contesto di una più generale visione dell'evoluzione della disciplina dai suoi primordi allo stato attuale.

A un altro illustre protagonista di questa storia, Dino Buzzetti, dobbiamo il contributo *Che cos'è, oggi, l'informatica umanistica? L'impatto della tecnologia*, che propone una riflessione dal notevole spessore teorico sulle modalità di interazione degli strumenti di elaborazione e comunicazione digitali con i fondamenti metodologici dell'informatica umanistica, che hanno portato paradossalmente portato al prevalere di iniziative di ricerca povere di contenuto teorico, ancorché ricche dal punto di vista della quantità (come le grandi iniziative di digitalizzazione testuale dell'ultimo decennio); il saggio si conclude indicando, come possibile linea di sviluppo

futuro, il recupero di una metodologia di ricerca fondata sulla elaborazione semantica.

Nel volume non trova espressione invece la bella tavola rotonda che, insieme agli autori, ha visto protagonisti Anna Maria Tammaro e Stefano Tortorella³ e in cui si sono dibattute le prospettive disciplinari e istituzionali dell'informatica umanistica, anche alla luce delle recenti riforme del sistema universitario, senza nascondere le criticità nonché le carenze e gli errori compiuti nel passato dalla comunità degli studiosi.

La seconda sezione, intitolata *Il testo e il computer*, prende spunto dall'omonimo libro di Giuseppe Gigliozi⁴. I temi affrontati nel libro andavano dall'esposizione delle basi concettuali dell'informatica agli strumenti più utili per l'umanista, passando per gli ipertesti, le biblioteche digitali e la modellizzazione delle strutture narrative. Con l'ironia e l'innata capacità didattica che lo distingueva, Gigliozi ha insomma affrontato nella sua carriera l'intero ventaglio delle questioni, dei problemi e delle potenzialità ancora al centro del dibattito sull'informatica umanistica: che cosa significa digitalizzare un testo letterario? A che cosa serve l'analisi informatica dei testi? Qual è lo stato dell'arte delle biblioteche digitali? Quali sono le sfide future dell'analisi e della gestione dei testi in rete? Con tali domande

³ Docente di Archeologia e membro per l'area linguistico-letteraria presso il CUN, nonché amico di Giuseppe Gigliozi.

⁴ Pubblicato per la prima volta a Roma nel 1993, in una prima versione con il titolo *Letteratura, modelli e computer*, il testo è stato poi edito in prima edizione da Bruno Mondadori col titolo *Il testo e il computer. Manuale di informatica per gli studi letterari* nel 1997, e in seconda edizione, con la revisione e la cura di Fabio Ciotti, e il titolo *Introduzione all'uso del computer negli studi letterari*, nel 2003, dopo la scomparsa di Gigliozi.

si sono misurati, dai rispettivi punti di vista e partendo dalle singole esperienze di ricerca, gli interventi raccolti in questa sezione.

Maurizio Lana nel suo *Da una digital library del latino tardo ad un corpus globale* descrive un importante progetto in corso presso l'Università del Piemonte Orientale per la realizzazione di una *digital library* destinata a raccogliere i testi latini tardi, basata sullo standard XML/TEI, e l'idea ambiziosa di procedere a una successiva cooperazione e integrazione con gli altri grandi progetti di archivi testuali della latinità.

Paolo Mastandrea, con il saggio *1991-2011, un ventennio di evoluzione nella filologia pratica*, si pone invece su un piano meta-teorico (ma senza cedere alle lusinghe dell'iper-specialismo), riflettendo sugli indubitabili vantaggi che le tecnologie digitali hanno portato al lavoro filologico, e mostrando come i presunti difetti da molti ascritti alla testualità digitale, siano in realtà il frutto di meccanismi operanti da sempre nella tradizione dei testi.

Sullo stesso ordine di discorso si pone il saggio di Raul Mordenti, *Domande teoriche sul concetto di edizione (nel nome di Giuseppe Gigliozzi)*, che propone un radicale ripensamento della forma libro in generale e, in particolare, di quel genere di libro che ha avuto un ruolo centrale negli studi testuali, l'edizione critica. A partire dalla considerazione che ogni innovazione tecnologica si afferma socialmente e culturalmente attraverso due fasi, la prima conservativa, in quanto cerca di preservare forme e contesti del medium precedente, e la seconda trasformativa, che afferma le potenzialità implicite e irriducibile del nuovo medium, Mordenti suggerisce come sia ormai il momento di liberarsi delle sovrastrutture dell'epistemologia gutenberghiana, come il ruolo dell'editore nella filiera del libro, o il concetto di testo ne *varietur*.

Il saggio di Federica Perazzini, *Oltre i confini del testo: le Digital Humanities tra scienza e opportunità*, fornisce invece un prezioso resoconto (basato sulla sua esperienza di ricerca presso lo *Stanford Literary Lab*) su una recente svolta metodologica che le tecnologie informatiche hanno introdotto nello studio del patrimonio testuale e culturale: l'applicazione degli strumenti di *text mining* e *knowledge extraction* per l'analisi computazionale di vasti archivi testuali/documentali e banche di dati.

Infine, Francesca Tomasi con il suo saggio *Digital editions between embedded markup and external representation. A case study: Vespasiano da Bisticci's Letters* offre un esempio concreto (applicato all'epistolario di Vespasiano da Bisticci) di come gli sviluppi recenti nelle tecnologie del web possano aprire la strada a forme innovative di edizione digitale di testi e *corpora*, in grado di esplicitare diversi livelli di contenuti interni e di collegarli con risorse esterne al testo: «the purpose of the digital edition is on one hand to represent the information that is implicitly connected inside the source... on the other hand to create semantic links with WWW resources, useful in order to describe this information».

La terza e ultima sezione si pone in un orizzonte più prospettico e strategico, come testimonia il titolo *Oltre l'informatica umanistica: culture digitali e reti della conoscenza*. Si cerca qui di riflettere sul ruolo dell'informatica umanistica nel panorama attuale delle scienze umane e sul significato che si attribuisce alle nuove "culture digitali", anche in relazione ai temi della formazione. "Informatica umanistica" è termine che ha resistito all'usura e alle perplessità per oltre venti anni; non si può dire lo stesso nel campo internazionale, dove da *Humanities Computing* si è passati al più spendibile ma anche più vago *Digital Humanities*, allargando il campo della riflessione e delle

ricerche a settori prima distanti o marginali. Ma la digitalizzazione non ha cambiato solo le pratiche di ricerca e innovato le sue metodologie. Essa ha ormai investito ogni campo della cultura, realizzando sul campo l'interdisciplinarietà auspicata dai suoi primi pionieri e ponendo oggi sfide complesse che riguardano non più solo l'accesso, la gestione e la rappresentazione delle forme culturali, ma l'assetto stesso delle nostre identità.

I lavori inclusi in questa sezione rappresentano dei tentativi, parziali e non sistematici, ma assai stimolanti, di predisporre una mappa di questo territorio proteico.

Domenico Fiormente nel suo *Per una critica culturale delle Digital Humanities*, fornisce forse il contributo più provocante della silloge, proponendo una critica culturale e politica delle Digital Humanities intesa come critica dei suoi presupposti ideologici e culturali taciti, a partire dal suo orientamento geopolitico, concludendo che: «l'implicito appiattimento sulle politiche tecnologiche, commerciali e industriali da un lato e l'origine tendenzialmente monoculturale delle rappresentazioni simboliche e logiche sono [...] fra i principali ostacoli per un'espansione non meramente strumentale delle DH». La via di uscita che egli propone consiste in una rinnovata contaminazione interdisciplinare e interculturale con le scienze sociali e antropologiche.

Gli estensori di questa introduzione si collocano invece in una ottica progressista, per così dire, e propongono di esplorare la strada delle recenti tecnologie del web semantico e dei Linked Data, al fine di ripensare prassi e strumenti delle Digital Humanities e dare un nuovo senso alla grande quantità di prodotti e risorse digitali creati in questo ultimo ventennio: Fabio Ciotti nel suo *Web semantico, linked data e studi letterari: verso una nuova convergenza*, si concentra in particolare sul dominio degli studi letterari; Gianfranco Crupi, con

Universo bibliografico e semantic web, rivolge invece la sua attenzione sul mondo delle scienze del testo e del documento, da lungo tempo compagne di strada delle Digital Humanities e delle scienze umane in generale.

Il contributo di Emiliano Ilardi, *La narrazione in rete dei patrimoni culturali digitalizzati*, introduce una ulteriore linea di riflessione. Rovesciando infatti la vulgata postmodernista, che vede nei nuovi media digitali l'affermarsi di un modello reticolare e decentralizzato della comunicazione/interazione sociale, Ilardi osserva come il successo di strumenti/ambienti quali Facebook, Twitter o Youtube sia dovuto non solo e tanto al fatto che essi permettano agli utenti di produrre e disseminare contenuti, quanto alla possibilità che forniscono loro di *raccontare* e *raccontarsi*. La narrazione è una strategia comunicativa e di senso che unisce mass media e new media e — propone l'autore — solo adottando strategie adeguate di *digital storytelling* le istituzioni culturali e della memoria potranno giocare un ruolo da protagonista nel nuovo ecosistema digitale.

Christoph Bläsi, nel suo *The book as a medium: interesting phenomena to be researched with the help of digital humanities approaches*, mostra come i metodi delle Digital Humanities non debbano limitarsi allo studio del testo, inteso come oggetto semiotico astratto, ma possono fornire importanti contributi anche per lo studio del libro come artefatto culturale, nella prospettiva dei Cultural Studies. Ne è testimonianza un importante programma di ricerca condotto presso la Johannes Gutenberg University di Mainz (dove è stata sviluppata un'innovativa infrastruttura digitale di supporto alla ricerca il VRE – Virtual Research Environment), volto a studiare la catena del valore nell'economia del libro e i processi di convergenza dei media.

Infine, Stefano Lariccia nel suo contributo, *Automi e linguaggio nell'ecosistema delle reti digitali*, fa una rassegna di alcuni strumenti disponibili nell'ambito del Natural Language Processing (NLP) e delle applicazioni, anche a livello internazionale, di linguistica computazionale.

La grande varietà di approcci teorici e disciplinari, di contenuti e di stili argomentativi dei lavori del convegno e dei saggi qui pubblicati testimoniano della vitalità di un'area di studi, le Digital Humanities, i cui confini sono ampi e indefiniti ma che, secondo molti, costituirà *the next big thing* nella storia delle scienze umane. Alla costruzione di questa impresa intellettuale, Giuseppe Gigliozzi, pur dalla periferia dell'impero, ha dato un grande contributo e, di questo, la comunità scientifica gli è perennemente grata.

Ringraziamenti

Desideriamo ringraziare tutti coloro che hanno contribuito a vario titolo alla organizzazione del convegno, alla sua riuscita e alla pubblicazione di questi atti. Per non far torto a nessuno li elenchiamo qui in ordine alfabetico: Rino Caputo, Domenico Fiormonte, Giovanni Ragone, Gino Roncaglia, Anna Maria Tamaro.

Un grazie speciale va infine a Cristina, Ginevra e Gianandrea Gigliozzi.

Fabio Ciotti

Gianfranco Crupi

A Giuseppe, amico e maestro

Conobbi Giuseppe più di venti anni fa quando per la prima volta bussai alla porta del suo studio, per chiedere la tesi di laurea. Dentro quello studio c'erano tre computer con gli schermi verdi e neri, e un professore gentile e disponibile che mi accolse con un sorriso, mi chiese subito di dargli del tu e mi contagiò con il suo entusiasmo per quel campo di studi in cui il testo si incontra con il computer. Di questo campo di studi Giuseppe era stato uno dei pionieri sin dagli anni ottanta, quando faceva parte del gruppo di studiosi di Informatica Umanistica che si riunirono intorno a Tito Orlandi.

Pochi anni dopo, insieme a me e ad alcuni altri amici e collaboratori, fondò il CRILet, Centro Ricerche Informatica e Letteratura, presso l'allora Dipartimento di Italianistica dell'Università "La Sapienza". Ricordo ancora oggi il pomeriggio in cui, dopo una sterminata serie di tentativi, riuscimmo finalmente a trovare un acronimo che non somigliasse a un personaggio delle *Cosmicomiche* di Calvino. In quegli anni ormai lontani di Internet, multimedialità e rivoluzione digitale in Italia non parlava ancora nessuno. Occuparsi di computer e letteratura era un vero atto di coraggio intellettuale. E questo coraggio era una delle caratteristiche di Giuseppe. Un coraggio che lo ha sostenuto per tutti i lunghi anni in cui il suo lavoro di studioso non è stato sufficiente a fargli ottenere quel riconoscimento accademico che meritava.

Ma il coraggio intellettuale di Giuseppe era intimamente legato alla sua intelligenza e alla profonda riflessione sul senso delle

ricerche a cui aveva dedicato la sua vita. Era convinto che l'incontro tra il mondo dei saperi umanistici e le tecnologie avrebbe cambiato il modo e i metodi con cui entrambe le discipline trattano i loro oggetti di studio. Quando il testo e il computer si incontrano nasce qualcosa di nuovo. Su questo cambiamento Giuseppe ha riflettuto e ha sempre chiamato a riflettere i suoi allievi. Credo che oggi nessuno può negare quanto questa sua convinzione rispondesse a verità.

Nel perseguire questa linea di ricerca, tuttavia, ha sempre mantenuto una onestà intellettuale e un rigore scientifico altissimo. Non ha mai ceduto al semplicismo o alla moda culturale come molti che in quegli anni saltarono opportunisticamente sul carro della multimedialità e del digitale. Questo rigore scientifico si trova in tutti i suoi scritti, in tutte le ricerche che ha coordinato e a cui ha partecipato. E soprattutto si ritrova nell'insegnamento e nei consigli che ci dava giorno per giorno nei lunghi e bellissimi anni che abbiamo passato insieme.

Ma sopra ogni altra cosa ciò che ricordo ancora di Giuseppe sono la sua disponibilità la sua gentilezza e il suo sorriso. Queste caratteristiche lo rendevano qualcosa di più che un professore e uno studioso: lo rendevano un maestro.

Con il gruppo di giovani studiosi che si erano raccolti intorno a lui, di cui facevo parte insieme a Domenico Fiormonte, Myriam Trevisan, e altri le cui storie di vita ho perso di vista, discuteva, trattandoci sempre da pari, e accoglieva senza nessuna presunzione le idee e le critiche. Condivideva tutte le sue riflessioni e ogni volta che doveva scrivere qualcosa o partecipare a un convegno chiedeva a me e agli altri cosa pensavamo. E naturalmente lo stesso valeva per tutti i nostri lavori. Non di rado ci trovavamo a pensare esattamente

la stessa cosa, e persino a dirla senza esserci consultati. Insomma Giuseppe era riuscito a creare una vera e propria scuola. E questa scuola è la sua eredità più importante.

Del rapporto di amicizia personale che ebbi con lui e con la sua famiglia, dell'affetto che ci legava, oggi a distanza di dici anni dalla sua scomparsa, fa ancora male parlare. I ricordi si affastellano nella mente, ma troppo vividi sono ancora quelli di quei tristi e assurdi ultimi mesi.

Ma questa disponibilità, questa capacità di ascoltare Giuseppe la donava non solo a chi gli era più vicino: era per ogni suo studente, per ogni collega (e non poche furono le delusioni che dovette patire, specie da questi ultimi). In dieci anni di lavoro insieme ho visto crescere intorno a lui tantissimi giovani, molti dei quali grazie al suo magistero, hanno avuto la possibilità di realizzare cose importanti dentro e fuori dall'Università. Per tutti Giuseppe aveva la stessa attenzione, lo stesso garbo, la stessa pazienza. E tutti lo hanno ricambiato con un grandissimo affetto e gratitudine.

Per tutti Giuseppe aveva quel sorriso che più di ogni altra ancora oggi mi manca infinitamente.

Fabio

Lo statuto dell'informatica umanistica

Introduzione ai lavori

Giovanni Ragone, Sapienza Università di Roma

Care amiche e amici, care colleghe e colleghi,

spero che questa iniziativa segni il rilancio di una tradizione di ricerca che ha avuto in Sapienza radici forti, per merito di due studiosi come Giuseppe Gigliozzi e Tito Orlandi, con i molti altri che hanno collaborato con loro. Una storia che non è andata dispersa. Potremo anzi riprendere il cammino, se sapremo accogliere energie di studiosi esperti e di studiosi giovani, costituendo una sezione scientifica nel nuovo ambiente del Centro di ricerca e servizi DigiLab. La nostra ispirazione parte dalla consapevolezza delle trasformazioni radicali della società delle reti e dunque della necessità ineludibile di un lavoro di sviluppo e riconversione del tessuto stesso delle scienze umanistiche nell'ambiente digitale. DigiLab ha assunto come terreno di azione il *Digital Heritage*, a partire dalla creazione, conservazione e comunicazione on line dei patrimoni e delle risorse culturali. La sua missione fondamentale è duplice: creare una grande Digital Library universitaria, aperta e pronta a confluire in una rete universitaria e non universitaria che renda fruibile e ri/mediabili le risorse; ma allo stesso tempo

accompagnare e favorire la mutazione che le nostre università dovranno affrontare nella prossima fase. Esse diventeranno strutture ibride con le reti: università delle reti. Avremo problemi anche drammatici da affrontare nel funzionamento, nell'organizzazione, nella struttura complessiva delle attività di formazione, ricerca fondamentale, ricerca applicata e servizio; e una simile mutazione non potrà avvenire senza che nelle università si sviluppino sul terreno dell'innovazione tecnologica e metodologica, della sperimentazione negli ambienti digitali, dell'apertura verso le reti, della collaborazione tra loro e con molti altri soggetti. Per questo DigiLab sperimenta innovazioni e servizi non solo riguardo alla infrastruttura della Digital Library e della comunicazione dei patrimoni, ma anche nei settori dell'editoria digitale, dell'e-learning, delle produzioni web e video, del 3D e delle installazioni digitali anche locali.

È in questo ambito si colloca anche lo sviluppo teorico e applicato della ricerca sui due versanti di cui si discute nel nostro convegno: quello delle scienze del testo nella loro ibridazione con l'informatica e quello più complessivo delle culture digitali, soprattutto per quanto riguarda il campo umanistico. La sezione di *digital humanities* di DigiLab avrà autonomia e sarà costituita da una rete di ricercatori della Sapienza e di altre università o centri di ricerca. Chiederemo a chi ne farà parte di darsi un'organizzazione formale, un programma di lavoro credibile, e obiettivi definiti. Ci attendiamo ricerca avanzata, collaborazione a livello nazionale e internazionale, e idee per corsi di formazione. Sosterremo, se il progetto troverà energie per camminare, una intensa attività di

workshop, da mettere a disposizione in video per tutta la comunità; ospiteremo le attività di laboratorio e la biblioteca di informatica umanistica.

Alle scienze del testo sono dedicate in questo convegno diverse sessioni, a cui partecipano colleghi molto più esperti di me. Il mio contributo riguarda invece lo scenario di fondo, a cui tutti facciamo riferimento, e sul quale mi esercito come direttore di DigiLab e come mediologo: il *digital heritage*, il lavoro delle istituzioni culturali, e il cambiamento necessario delle università¹.

Adattarsi, creativamente, e creare hub o perire. Dove va il lavoro culturale in un ambiente digitale?

Le mappe che proverò a tracciare riguardano il futuro delle istituzioni culturali. Una grande profezia ha sostenuto nell'ultimo Novecento la visione democratica sulla formazione, sulla cultura, sul lavoro: quelle sulla "società della conoscenza", e su una formazione resa accessibile a tutti ovunque e per tutta la vita dalle tecnologie digitali. Profezia fondata su processi reali, e capace di mobilitare politiche, ceti intellettuali, movimenti sociali. Ora essa è in parte svanita, ma non perché la situazione sia ferma ai paradigmi della società industriale. La profezia svanisce davanti all'evidenza dei nuovi conflitti e di una compressione dei diritti garantiti dalle

¹ Una versione più estesa di alcune linee qui esposte è nell'articolo di G. Ragone, A. Ceccherelli, E. Ilardi, "L'università delle reti". In *Scuola Democratica*, 2011(3), pp. 91-114; e per una teoria completa cfr. "Digital heritage: memoria, cultura, tecnologia e istituzioni ibride". In D. Capaldi, E. Ilardi, G. Ragone (ed.), *I cantieri della memoria. Digital Heritage e istituzioni culturali*, Napoli, Liguori, 2011.

vecchie forme democratiche.. Mentre le dinamiche e economiche e sociali del modello di sviluppo basato sulla produzione di informazione richiedono sempre più competenze di tipo intellettuale, trasformando il lavoro manuale in lavoro anche intellettuale, la pressione della globalizzazione e le stesse dinamiche evolutive della comunicazione determinano nuovi sistemi di sfruttamento e di dominio, nuove diseguaglianze sia nella gestione della conoscenza che nell'accessibilità reale della formazione. La velocità del cambiamento è tale, che le istituzioni culturali, destinate come le istituzioni politiche a garantire - attraverso patti culturali e sociali - livelli determinati di accesso alla cultura e alla formazione, non riescono a tenere il passo, e sembrano aver perduto la capacità di interpretare questa fase ormai compiuta di transizione verso la "società in rete", e di definirne i modelli.

Nel caso italiano viviamo un vero e proprio sbandamento; le istituzioni e le università reagiscono su se stesse, più che aprirsi al contesto culturale e sociale. Tanto che le ultime vicende relative a riforme, norme e tagli di bilancio in tutto il settore della cultura e della formazione, con i conseguenti riflessi sulla qualità, hanno in molti sensi obliterato nella discussione quello che dovrebbe essere considerato il terreno reale. Proviamo a rendere esplicita la *vera* e per certi versi drammatica domanda che si tende a rimuovere: se la società è in rete, ci saranno e come saranno istituzioni culturali, formazione, e - per quanto riguarda le riflessioni che seguono - una università in rete? Su questo, non si riflette abbastanza, e quindi non si governa il cambiamento.

Gli umanisti lavorano sull'eredità e sul patrimonio (*heritage*) per comprendere i processi della cultura, e anche per attrezzare l'umanità ai colpi del cambiamento dell'ambiente. L'*heritage* contribuisce alla nostra identità in modo determinante; è un complesso di relazioni sociali, valori, e pratiche di solito ben definite e specializzate: si tratta sia di azioni individuali e professionali, socialmente legittimate, come quelle dell'archeologo, del restauratore, dell'esperto d'arte, collezionista, mecenate o mercante, ma anche dell'aedo e del narratore, dell'insegnante, dell'urbanista o progettista che interviene sui territori o sui sistemi turistici; sia di azioni organizzate di vere e proprie istituzioni come archivi, biblioteche, musei – ma anche università, accademie, fondazioni, e altro ancora. La struttura e la *forma estetica* (comunicativa) delle azioni individuali e delle stesse istituzioni ha origini storiche, determinandosi nello scambio e nella tensione tra una domanda organizzata e un'offerta (di informazione, di formazione, ma anche di ospitalità e di esperienza immaginaria, piacevole o perturbante), entrambe modellate secondo processi identitari variabilmente aperti e flessibili. La scuola pubblica di massa, per esempio, assume la sua forma tradizionale nel rapporto tra domanda di istruzione di base generalizzata e opportunità, colta dagli stati nazionali, di formare cittadini-lavoratori e secondariamente le donne, secondo valori, standard e immaginari omogenei nella trasmissione del sapere da una generazione all'altra.

Ma oggi *chi* rivolge le sue domande verso le istituzioni dell'*heritage*? E *per quali domande e servizi*? Le ricerche in questo campo si autolimitano di solito a una dimensione empirica e

descrittiva: i musei, le fondazioni e gli enti, analizzati soprattutto secondo parametri economici; gli "utenti" nella loro composizione o tipologia; l'evoluzione normativa; la situazione dei diversi paesi... Talmente centrale per la nostra vita individuale e collettiva è il complesso di funzioni sociali che trasmette i patrimoni e le conoscenze da rendere in un certo senso familiare e scontato un intero paesaggio culturale, dove sono principalmente le istituzioni a dover assicurare la conservazione e la mediazione dei "beni" verso il "pubblico". L'istituzione che offre un servizio, lo offre in teoria in funzione dei suoi impliciti utenti. O almeno così dovrebbe essere, se restiamo fedeli all'idea di una società ordinata, o per virtù del mercato, o perché generosamente aperta ai bisogni dei cittadini. Oggi però tutto questo è in crisi, e lo è perché le istituzioni devono adattarsi al nuovo ambiente mediale, dove le conoscenze e i patrimoni saranno comunicati e diventeranno strumento per apprendere, memoria vivente. Si ripete ciò che è più volte accaduto dopo ogni rivoluzione dei media, dell'economia e delle infrastrutture: le conoscenze, i patrimoni, l'ambiente di ricerca e di apprendimento devono essere accessibili nei *nuovi* modi della comunicazione e della mente che si sono ormai assestati. Non solo: per essere di continuo legittimata nella sua credibilità, fungibilità e autorevolezza, l'istituzione deve essere riconoscibile e fungibile nella sua specificità anche nel nuovo ambiente. E se questo non accade, l'istituzione è destinata a scomparire.

Oggi conoscenze e patrimoni vanno resi accessibili nello sterminato oceano del web. Tanto che le diverse funzioni, prima distinte, delle università, sono sempre più intrecciate dovendo

rimodellarsi secondo il nuovo paradigma, antitetico a ogni distinzione e votato alla disponibilità "orizzontale" e ibrida di ogni genere di operazione. Si può ragionare seriamente, per esempio, di e-learning, formazione continua, apprendimento collaborativo solo se l'università intende rendere disponibili on line i documenti e le ricerche che essa produce, i suoi archivi, il suo sapere, le sue community: se intende rendere accessibile se stessa in quanto struttura, o forse oggi sarebbe meglio dire infrastruttura.

L'utopia di fine Novecento vedeva nel passaggio alla *network society* l'avvento della democrazia reale e compiuta in tutti gli ambiti della società; e secondo questo modo di pensare le relazioni sociali, le informazioni, le conoscenze sarebbero in futuro autoprodotte e scambiate in un'infrastruttura totalmente orizzontale, che eliminerebbe utopisticamente la mediazione di qualsivoglia centro di potere o fonte privilegiata di autorità. Costruire una rete delle conoscenze democraticamente distribuita avrebbe reso obsolete le possibilità di *gatekeeping* informativo o formativo. Ciascuno sarebbe in grado di selezionare del tutto autonomamente le informazioni senza costrizioni o controlli dall'esterno e di riorganizzarle in un personalissimo *agenda setting*. Ma la teoria ha dovuto registrare all'inizio del nuovo millennio il fatto che l'evoluzione del web procede in modo diverso. Internet viene descritta come una rete sì, ma sempre più evidentemente organizzata intorno a pochi e potenti nodi (infra)strutturali, che regolano, organizzano e oggi sempre di più controllano o limitano il flusso di informazioni. Sono i grandi motori di ricerca, i principali *social network*, i grandi siti di commercio on line e alcuni blog specializzati ad accaparrarsi la maggior parte

degli accessi e dei link: giganteschi connettori ai quali non si può più sfuggire, che filtrano l'immenso flusso di bit e decidono del grado di visibilità delle informazioni che circolano. Con la crescita di una sterminata massa di informazione, i momenti di selezione diventano ancora più importanti, perché separano il poco che verrà letto dall'oceano dei contenuti sommersi. Soprattutto i provider e i motori di ricerca esercitano un controllo, non con l'esclusione delle notizie, ma attraverso la loro allocazione in zone periferiche e buie, quasi impossibili da raggiungere (questo, più o meno, sostiene la *gatekeeping theory*)². I contenuti dei *prosumer* – della cosiddetta "autocomunicazione di massa" – risultano invisibili se non passano per i cosiddetti *hub* della rete. Non solo: essi vengono scartati se non ne rispettano i *frame* narrativi.

È in questo contesto ineludibile che sono costrette a giocare il loro ruolo le vecchie istituzioni culturali (scuole, università, musei, istituti culturali, archivi). E il processo è probabilmente decisivo per il controllo delle sorgenti narrative collettive. Riusciranno le istituzioni a essere ancora attori fondamentali? La risposta può essere positiva, ma solo se esse decidono finalmente di diventare connettori e nodi della rete, e in definitiva di aggregarsi in *hub*. In Internet, ormai è chiaro, la quantità di contenuti digitali che i nodi gestiscono produce la forza di gravità necessaria ad attrarre gli altri

² A. Miconi, *Reti. Origini e struttura della network society*, Roma-Bari, Laterza, 2011, p. 95. Su questi temi vedi anche: N. Carr, *Il lato oscuro della rete. Libertà sicurezza, privacy*, Milano, Rizzoli-ETAS, 2008; C. Formenti, *Cybersoviet. Utopie post-democratiche e nuovi media*, Milano, Raffaello Cortina, 2008; G. Lovink, *Zero comments. Teoria critica di internet*, Milano, Bruno Mondadori, 2007.

contenuti. Le università italiane, per esempio, possono contare, potenzialmente, tra forze interne e relazioni con community esterne, su almeno due o tre milioni di *prosumers*, una massa che interagisce in modo significativo con quella "classe creativa" – il 30% circa degli occupati – che costituisce la riserva residua di energia del nostro paese. Non è poco. Il passo necessario da compiere è dunque probabilmente quello delle grandi *digital libraries* universitarie, disponibili in modalità il più possibile *open access*. La teoria della "coda lunga" di Anderson vale anche per l'università: per valorizzare tutti i contenuti, anche i più specialistici o di nicchia (come fa ad esempio Amazon), e poi per attirare i contenuti prodotti dagli utenti (come fanno YouTube o Facebook), c'è prima bisogno che si formi un campo di attrazione di un grande nodo di rete³. Ovvio che poi a parità di informazioni a fare la differenza sarà la *qualità* delle conoscenze e della comunicazione. Anche in termini di *attenzione*⁴.

³ C. Anderson, *La coda lunga. Da un mercato di massa a una massa di mercati*, Torino, Codice, 2007.

⁴ "[...] il valore di uno scambio comunicativo è connesso con il grado di novità e con l'intensità emotiva che provoca. L'economia della conoscenza viene da questo punto di vista a coincidere con l'economia dell'attenzione: la capacità di produrre novità che vincano nella competizione per catturare la vera risorsa scarsa nel mercato, l'attenzione delle persone. La società della conoscenza è quindi sostanzialmente regolata da due forze tipicamente umane, le stesse che rendono utile la tecnologia che la incarna: la capacità di produrre novità, ovvero il capitale intellettuale e la capacità di catturare l'attenzione, ovvero il capitale sociale." (S. Bagnara, M. Mesenzani, "Lavoro e formazione nell'età della conoscenza", in *Scuola Democratica*, nuova serie, 2010(1), p.106).

Su tutto questo investono le migliori università americane ed europee.

In questa fase storica, una gran parte delle azioni di cambiamento istituzionale delle università e della formazione superiore potrebbe essere etichettata come un tentativo di sopravvivere cambiando il meno possibile in condizioni sostanzialmente mutate. In Italia siamo particolarmente abituati a operazioni di pura cosmesi, come stanno diventando i tentativi di regolazione burocratica misti alla retorica della "qualità" e della valutazione; e comunque negli ultimi tempi la credibilità delle azioni relative è crollata, perché ciò che si afferma come obiettivo delle riforme stride con l'anarchico affidamento di attività strategiche come la ricerca e la formazione al sostegno di "mercati" sempre più deboli per il contrarsi della capacità economica delle famiglie e delle imprese, sommato al taglio del welfare e dell'istruzione pubblica. Siamo ancora di fronte ad iniziative residuali, che innescano dinamiche per l'ennesima volta prevalentemente autoreferenziali, a fronte di scarsi risparmi e scarsi miglioramenti. In un certo senso, sono gli ultimi fuochi di vecchie politiche, sia di sinistra che di destra, che si affidano – senza comprendere i processi reali – o all'illusione di una mano statale che imponga ristrutturazioni istituzionali qualitativamente apprezzabili, o all'illusione che la mano statale sia sostituibile da quella di un mercato nazionale in grado di produrre di per sé sviluppo delle attività culturali, di ricerca e di formazione. E sempre più spesso, ad entrambe le illusioni insieme.

Chi volesse provare sul serio a cambiare passo dovrebbe capire che non ha senso declinare iniziative pubbliche senza che siano

chiare le direzioni del cambiamento globale, nella loro dimensione più imponente, radicale e strutturale: occorrerebbe insomma che avesse chiarezza sul *come* si determina il movimento generale della cultura, dell'economia e dell'università delle reti.

In generale, l'apprendimento e la ricerca sul web, attuandosi fondamentalmente "fuori dai luoghi" materiali e mostrandosi altrettanto o più flessibile delle forme tradizionali e perfino in grado di integrarsi con queste in nuove forme ibride - mettono ovviamente già di per sé in questione le attuali istituzioni. Negli Stati Uniti, in Europa, in Oriente, in America Latina, pur tra profonde differenze, lo sviluppo economico, sociale e culturale si accompagna negli ultimi anni a una sostanziosa crescita e ristrutturazione della rete universitaria, dei suoi nodi e delle sue forme, della sua accessibilità sul web; e a un intensivo sviluppo dell'e-learning. Non solo il Nord America e l'Europa, ma il resto del mondo sta incamminandosi strategicamente verso l'e-learning, sia come alternativa all'aula sia nelle forme *blended*: nel 2014 l'Asia, continuando con un tasso di crescita annuo del 33,5%, supererà il nostro continente, e uno dei fattori di successo è l'opinione che nei sistemi e-learning vi sia una migliore accessibilità delle informazioni e delle conoscenze, e la possibilità di apprendere con docenti più "famosi"⁵. Ma ciò va accettato come parte di un cambiamento molto più complesso. *Tutte* le istituzioni, infatti, sono destinate a una veloce metamorfosi,

⁵ J. Baggaley, T. Belawati, J. Daniel, *Distance Education Technologies in Asia*, New Delhi, Sage Publications, 2010.

avviandosi verso un mondo dove gli uomini conducono la loro esistenza prevalentemente in ambienti digitali.

Il punto allora è: quali sono i caratteri specifici di tale metamorfosi riguardo alle pratiche culturali della ricerca scientifica e dell'*heritage* - le due funzioni principali delle università, ma anche di diverse altre istituzioni? È su questo che occorre mettere a fuoco la discussione.

Se osserviamo la deriva generale delle istituzioni culturali e scientifiche *nel* mondo digitale, alcune "domande" si presentano già con molta forza, e sono strategicamente ineludibili; ad esse le istituzioni della formazione e della ricerca sono chiamate a rispondere o decadere, entro una "forbice" sempre aperta tra le già elevatissime possibilità offerte dalle tecnologie e le ancora deboli attività di virtualizzazione. Le elenchiamo rapidamente, perché sono domande da tempo sul tappeto:

- la richiesta di ri-orientare strutturalmente l'attività verso il web e la ri/mediazione collaborativa (viene dalle ultime generazioni di studenti);
- il richiamo a basarsi sul cambiamento in corso dei "target" e degli attori, non a respingerlo (viene, per quanto debolmente in Italia, dalle imprese e dai servizi);
- il confronto su standard internazionali (viene comunque dalle autorità di regolazione economica e dai gestori del finanziamento europeo, e poi dalla competizione globale);
- la necessità di sviluppare capacità "federative" (viene, comunque, da esigenze di razionalizzazione dell'investimento e di ottimizzazione delle risorse);

- quella di fondere modelli "no profit" e attività di mercato (viene comunque dalla spinta all'autofinanziamento, nel diffondersi di una pluralità di modelli organizzativi e istituzionali con cui competere, in un mercato che è globale anche per l'università, la cultura, la ricerca).

Proviamo ora a esaminare lo scenario da tre punti di vista diversi (la virtualizzazione; la collaborazione; le missioni istituzionali relative al lavoro e all'innovazione).

La virtualizzazione e i suoi derivati

Virtualizzazione: con il XXI secolo l'industria culturale, le istituzioni culturali e gli attori individuali si sono integrati completamente nel cyberspazio digitale. Anche le università, come le fondazioni o i musei, poggiano su "contenuti" digitalizzati e su materiali direttamente prodotti in digitale. Il "salto" è sia quantitativo che qualitativo: la "convergenza" tra le reti, i sistemi della comunicazione, i media, ma anche il diffondersi a livello di massa di un *learning by doing* permettono a milioni di attori in Internet di diventare anche "produttori" di contenuti e di simboli in ambienti di interazione.

Il travaso "canonico" di riferimenti, esempi, classificazioni e norme dagli strati elevati della cultura a quelli intermedi e all'apprendimento scolastico, e da una generazione a quella successiva, non perdono certo del tutto il loro senso; le "canne" del canone continuano a essere infisse nel terreno; ma è un lavoro sia più violentemente esposto alla pressione massmediale che si esercita sugli spazi più "poveri" della persuasione politica - quelli soprattutto

della televisione generalista – sia più soggettivo e molecolare: si intensificano le pratiche di un ceto intellettuale diffuso e ora anche di ceti di istruzione recentissima, per declinare e costruire identità e storie: si pensi al prevalere – sotto i nostri occhi – di una storia "leggera" e multimediale, rispetto alla storiografia tradizionale: milioni di esperimenti storiografici divulgativi su decine di migliaia di temi, dalla pesca al tonno alla archeologia, dalla corsetteria al Giro d'Italia, destinati ad Internet, alla televisione, alla stampa, all'insegnamento, alla virtualizzazione istituzionale on line e ad allestimenti locali, senza escludere lo strumento del libro. Questo lavoro - in buona parte sfruttato dall'industria della comunicazione, ma destinato ad assumere un ruolo determinante anche nelle università - contribuisce a contrastare l'idrovora del presente assoluto, la violenta virtualizzazione di ogni spazio di vita nei format da *reality* e da spot pubblicitario, o la più morbida detemporalizzazione e deterritorializzazione dei social network. Occorre prendere consapevolezza dei cambiamenti che ciò comporta per istituzioni che continuino a svolgere un ruolo culturale fondamentale: esse diventano anche o soprattutto "macchine" mediali ibride per la virtualizzazione e la ri/mediazione.

Collaborazione. A questi caratteri si aggiunge la reticolarità. Tecnologie, disponibilità infinita di oggetti, e nuovi rapporti tra luoghi e flussi hanno cambiato il modo in cui si formano le identità collettive: il canone non è più fatto di opere e autori, ma di correnti culturali *mainstream*, basate sulle infrastrutture mediali e tecnologiche, dove moltitudini di attori si mettono in relazione con l'*heritage*, apprendono, proiettano immaginari, interagendo tra loro.

Rovesciando l'idea che la cultura e la scienza siano trasmesse dalle istituzioni a loro "destinatari", si può azzardare l'affermazione che oggi le stesse istituzioni culturali possono esercitare il loro ruolo proprio perché gli intellettuali del XXI secolo tendono, come e più degli altri strati, ad adottare pratiche collaborative in rete. E va notato che anche se dovessimo utilizzare i criteri di classificazione sociologica del passato, i ceti intellettuali rappresentano oggi in Italia almeno un 30% della popolazione: una moltitudine, per di più attiva su un mezzo così ampiamente aperto all'interazione con gli altri ceti sociali come è Internet. Non si tratta più solo delle *community* spontanee in rete, ma di un fenomeno collaborativo che pervade dall'interno le università e via via le altre istituzioni.

Nel mondo digitale, la virtualizzazione della cultura implica la collaborazione in rete, un procedere *wiki* che si basa su una apertura positiva alle competenze di ogni interlocutore effettivo o potenziale, sia su un cambiamento radicale delle procedure della conoscenza (l'attenuarsi delle attività di ordine classificatorio e deduttivo, e di converso l'accentuarsi delle attività di tipo associativo e analogico, semiconscie), sia su un ibridarsi delle strutture più stabili dei saperi con le facoltà immaginative (a livello collettivo le virtualizzazioni vengono "fissate" in testi o oggetti *mainstream*; e a livello individuale in configurazioni più stabili della memoria come connessioni neuronali). E questo paradigma reticolare apre nuove prospettive anche per le funzioni tradizionalmente ricoperte dalle istituzioni culturali, su ogni versante (conservazione della memoria, accessibilità, apprendimento, esposizione). La raccolta ed esposizione di oggetti e testi- individualmente o in piccoli ambiti e

reti settoriali considerati degni di considerazione, di interpretazione, di conversazione - genera piccoli microcosmi che si collegano ad altri. La selezione avviene *bottom up*, dal basso e in rete, ricercando ed elaborando individualmente, ma con l'intenzionalità di chi comunica, di chi lancia il primo messaggio e si apre alle controposte, non di chi si impone. E sta già avvenendo.

Il motore della cultura si è smaterializzato e diffuso in milioni di reti di produzione, che coinvolgono e riutilizzano miliardi di attori. Il baricentro si è spostato sulla ri/mediazione digitale e ha reso il XXI secolo differente da quello precedente per almeno tre motivi: a) la dimensione immensa, galattica del patrimonio e della produzione culturale e scientifica; b) la convergenza tra immagini fisse o in movimento, suoni e testi, sempre più "naturale", attraverso macchine ipertestuali "embedded" nei siti web e nei luoghi (tanto che non ha molto senso distinguere tra "realtà" e "virtualità", tra "presente" e "patrimonio", tra "soggetto" e "testo" o "comunicazione"), e c) la base connettiva della produzione, dove i sistemi tradizionali uno→molti (i massmedia) e io↔tu (la conversazione privata) convivono in un nuovo sistema reticolare e "neurale" di comunicazione, in cui ogni nodo (creato da un singolo o da un gruppo, impresa, istituzione) è potenzialmente accessibile e più meno disponibile a incrementare il suo "patrimonio" attraverso la connessione e collaborazione con gli altri nodi. Tutto questo, ossia l'iperaccelerazione della attività, la convergenza in un unico ambiente "virtuale", e la connettività, rende possibile una gigantesca attività di ri/mediazione nelle reti, che è inoltre efficace, verosimile, "naturale", diffusa e "democratica" nel

senso che non dipende strettamente dalle istituzioni culturali né dalla grande industria dell'immaginario⁶.

La delocalizzazione e detemporalizzazione tipiche della "formazione a distanza" assumono così ben altro senso (e si integrano con nuovi luoghi e tempi), trasformando i ruoli di gestione, e rendono la coppia "insegnamento"/"controllo" solo *una* delle molte varianti possibili. I "luoghi" della memoria, del patrimonio in rete e quelli dell'apprendimento diventano in larga misura sovrapponibili, o addirittura coincidenti. In effetti si gioca su questo terreno una delle più importanti partite della trasformazione dei modelli sociali e culturali: *learning organization*, centralità dell'apprendimento come fattore necessario per la sopravvivenza di ogni cultura ed organizzazione; tanto più necessario quanto più veloce è la trasformazione dell'ambiente. In ogni attività si può trasferire un modello organizzativo basato su quattro principi: conoscenze locali approfondite, apertura costante verso basi di conoscenza esterne, capacità di agire come comunità, trasferimento di ciò che i membri apprendono in nuove conoscenze.

L'apprendimento, in sintesi, è la base di un *design* innovativo che genera e determina tutte le altre attività di *knowledge management* finalizzate all'assolvimento di mandati sociali determinati. Ma si struttura anche come una comunità di pratiche, "mutual engaged": quello che per Castells è il fattore-chiave dell'inclusione/esclusione nella società delle reti, indispensabile per l'autocostruzione individuale e per la generazione di identità collettive, vale a dire

⁶*Ivi.*

l'acquisto delle competenze necessarie per garantirsi flessibilità, connettività, riprogettazione continua. In una cultura basata sulle soggettività, l'orientamento verso l'apprendimento è un altro tratto che caratterizzerà sempre più le diverse istituzioni culturali e scientifiche. Un "salto" evolutivo, che prevede di integrare nella struttura didattica, pedagogica e tecnologica dell'e-learning grosso modo tutti i processi tipici del cosiddetto Web 2.0, nel segno dei concetti-chiave definiti dal celebre articolo del 2005 di O'Reilly (*convergenza, partecipazione, condivisione...*)⁷. In coerenza con le teorie e le pratiche pedagogiche del socio-costruttivismo, che conoscono in questi ultimi anni una evidente ripresa e soprattutto numerosissime esperienze pratiche e sperimentazioni. Nel complesso, le pratiche dell'e-learning stanno ora gradualmente convergendo verso un insegnamento/apprendimento situato e basato sul lavoro di gruppo collaborativo o cooperativo, sulla condivisione di esperienze comuni, sulla partecipazione attiva (e quando possibile creativa) degli studenti nel processo didattico, e sull'attenzione per le possibilità di integrazione con l'apprendimento informale, che emerge proprio in virtù dello scambio relazionale tra i soggetti coinvolti. L'e-learning avanza in generale e nell'ambiente universitario e, come accade per altri fenomeni nella società in rete (si pensi per esempio alle reti dell'energia rinnovabile), è probabile che finisca per permeare *bottom up* i sistemi formativi. Anche per una ragione auto evidente: non è più immaginabile un futuro dell'educazione *anche* tradizionale senza

⁷ T. O'Reilly, *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*, 30. 09.2005, <http://oreilly.com/web2/archive/what-is-web-20.html>.

un ragionato ed intelligente utilizzo del web, sia come strumento di ricerca che come luogo di collaborazione e di condivisione, poiché le dinamiche relazionali e culturali delle nuove generazioni portano ad escludere ogni ipotesi di "resistenza" al nuovo medium. Le esigenze del *lifelong learning* vanno nello stesso senso, e con gli *smart device* si è compreso che il *mobile learning* sarà disponibile praticamente ovunque e sempre. L'università, per esistere, non può precludersi questo ampissimo spazio d'azione, sia in termini culturali che economici. Deve anzi necessariamente essere protagonista.

Innovazione e lavoro. La rivoluzione delle reti e i suoi derivati

Per ragionare sulle missioni sociali e del lavoro per le istituzioni e dell'università per la formazione e l'innovazione – occorre ripartire proprio da una rivisitazione critica dell'idea della "società della conoscenza". Uno degli effetti dello sbandamento attuale è una certa tendenza – vista la sovrapproduzione di laureati e di esperti rispetto al mercato del lavoro - a enfatizzare comunque una funzione indispensabile delle università nella creazione di *skill* di base, tra i quali le competenze di interpretazione delle conoscenze⁸. In effetti l'economia delle reti, il "modello di produzione informazionale" che produce informazione e a mezzo di informazione⁹, si presenta come un sistema in prevalenza

⁸ R. Moscati, "Le trasformazioni dell'università italiana". In *Scuola Democratica*, nuova serie, 1, 2010, pp. 94-98.

⁹ Cfr. M. Castells, *La nascita della società in rete*, Milano, EGEA, 2002.

terziarizzato, basato sull'invenzione e sulla gestione di servizi (finanziari, per le imprese, per le persone), che incorporano conoscenze, immaginari, tecnologie e – quando servono – luoghi e corpi. Non esiste un rapporto "lineare" tra (accumulo/scambio di) conoscenza e dinamiche di sviluppo (più conoscenze, più sviluppo); esiste piuttosto una *"osmosi" fra i processi culturali di produzione di conoscenze e immaginari e il sistema economico, condizionata dalle forme del dominio politico-economico* (le imprese e i servizi crescono sfruttando i processi culturali e mediali). Lo slogan della "società della conoscenza" continua in questo senso a rinviare a qualcosa di concreto, a una reale novità storica: l'emergere di una cultura e una economia in cui prevalgono l'immateriale sul materiale, l'informazionale sull'industriale, e una rivalorizzazione del lavoro intellettuale come comando, innovazione e gestione di ingenti strati di altri lavoratori intellettuali, o di lavoratori manuali a cui sono necessarie non irrilevanti conoscenze (si pensi a cosa è diventato il lavoro agricolo, per esempio).

Più che a *skill* di base, occorrerebbe riferirsi a una enorme crescita della varietà dei lavori e delle competenze utili, con profili che aumentano in modo esponenziale in un sistema nel quale interagiscono attori e livelli molto differenti: i grandi gruppi industriali a rete globale, con i loro apparati di produzione e di mercato, uno strato intermedio di imprese di produzione e di servizi a carattere artigianale, o istituzionale, variamente ibrido con le reti, e una continua germinazione profit e no profit di iniziative di comunicazione, direttamente in rete, più gli apparati politici e istituzionali. In generale, lo sbandamento delle nostre università

corrisponde alla difficoltà di adeguarsi al continuo aumentare e differenziarsi delle richieste che il lavoro e la vita sociale pongono agli individui e alle organizzazioni, in termini sia di conoscenze che di competenze (le conoscenze coniugate con l'abilità "pratica", operativa), ciò che ha comportato in tutto il mondo una continua ristrutturazione dei sistemi di istruzione e di formazione professionale. E va rimarcato che mentre quasi dovunque si assiste a un incremento degli investimenti pubblici e di quelli privati (che si dirigono verso i modelli ritenuti più efficienti), in Italia, dove l'aumento della domanda di istruzione superiore è in linea con le tendenze globali, la situazione degli investimenti è invece in controtendenza, con la profonda depressione che ne deriva. Il nostro scarso investimento, a sua volta, non è solo il risultato di scelte ideologiche. E' l'effetto in primo luogo dell'inefficienza strutturale del sistema-paese, che non riesce a ristrutturarsi in modo adeguato alle domande dell'economia globale delle reti (di conseguenza la centralità della ricerca e della formazione viene predicata solo a parole, e il potere politico può tagliare la spesa oltre misura senza provocare crisi di rigetto, o addirittura praticare scelte malthusiane, o favorire rendite di posizione particolari); e in secondo luogo dei modelli inadeguati delle istituzioni e delle università stesse, nonostante i tentativi di riforma (di conseguenza esse non attirano abbastanza investimenti, anche in proporzione alla modesta quota di risorse rese disponibili dallo stato e dal capitale privato per ricerca, sviluppo e formazione). Sia a livello di politiche pubbliche, sia operando nelle singole istituzioni, quindi, c'è spazio per ripensare e progettare.

Un secondo aspetto della crisi consiste nella discontinuità del sistema istituzionale e universitario per quanto riguarda il rapporto con le aree emergenti, ad alta densità connettiva. Nel nuovo modello di sviluppo le università sono utili solo se si intrecciano con quelle tipiche situazioni di addensamento di connessioni digitali, masse critiche di competenze qualificate, forti investimenti, creatività e capacità di organizzazione e di strategia globale/locale, che costituiscono oggi le pre-condizioni per l'ascesa economica e culturale di determinate aree del mondo. L'Europa, nel quadro attuale (o meglio: le aree europee che continuano a rispondere complessivamente bene allo sconvolgimento dei vecchi equilibri mondiali), sviluppa soprattutto produzioni *high tech*, adattate rapidamente alle domande di consumo dei paesi emergenti, e servizi alla persona; si tratta soprattutto delle aree del nord e centro Europa, favorite da un alto livello di istruzione, da infrastrutture di comunicazione avanzate, e da decenti apparati pubblici; dal basso debito pubblico ma soprattutto da una forte massa critica di attività di ricerca. Anche vaste aree del nostro paese, sufficientemente "dense" e interconnesse, sono incubatori della nuova economia; ma i fattori che rallentano o impediscono l'emersione, anche in quelle aree avanzate, sono gli scarsi investimenti, la separatezza degli attori, e l'inefficienza dei processi che favoriscono la loro collaborazione. Nonostante le posizioni "ufficiali" delle associazioni nazionali degli imprenditori, si pensi agli orientamenti spesso ostentati anche in pubblico di molti *leader* industriali, che negano l'utilità strategica delle università (ad esempio nel Veneto in crisi endemica da anni); si

pensi anche alla mancanza di coordinamento tra le università e gli altri attori pubblici, come i ministeri e le Regioni.

Un terzo aspetto critico riguarda il conflitto sui diritti di proprietà e di sfruttamento della conoscenza. I diversi livelli e soggetti della società delle reti hanno interessi contrastanti: le grandi *holding* puntano soprattutto sulla vendita di infrastrutture, hardware e servizi, e possono garantire un accesso apparentemente libero ai "contenuti", come si garantisce la possibilità di una "conversazione" privata su una connessione (in realtà i costi e un profitto sono addebitati attraverso il servizio, l'introito pubblicitario e la vendita dei profili dei consumatori). I soggetti intermedi, le imprese e i servizi "locali" (ma inseriti nei flussi globali), quando sono detentori di diritti di sfruttamento delle conoscenze puntano o a utilizzarli fino in fondo, difendendoli sul piano giuridico, oppure a negoziare con le *holding* un'integrazione come servizio nelle loro logiche, oppure ancora a sviluppare in proprio – nei limiti del possibile – servizi di accesso ai "contenuti" in rete. Ma i prezzi devono essere così contenuti e la qualità così elevata da competere con le *holding* che eventualmente decidessero di imitarli, e contemporaneamente da scoraggiare l'emulazione da parte di miliardi di *prosumer* e di milioni di community che potrebbero offrire servizi simili, secondo logiche completamente *no profit* o secondo logiche miste di autofinanziamento a costi ridotti, ma sempre contando su energie "collaborative", disponibili in buona parte nella logica del dono. Di qui la tendenza recente a integrare in qualche modo produttori "locali" e *prosumer*, in forme a volte perverse di *crowdsourcing* o in modi più convenienti per tutti. Come rilevante organizzazione "locale" che produce e

ri/media conoscenze, l'università come l'istituzione culturale si trova ad affrontare le stesse difficili condizioni dei soggetti intermedi.

Non è possibile quindi ridefinire la *mission* delle istituzioni e delle università senza ripartire dalle nuove forme dell'economia e del lavoro. Non è più sufficiente "coniugare" formazione di base, formazione professionale, apprendimento permanente come un solo sistema, e provare a renderlo più flessibile in vista della richiesta di aggiornamenti continui delle competenze individuali, a fronte di probabili cambiamenti di ruoli, mansioni, settori di attività... Questo cambiamento, che in Italia siamo lontani dall'aver realizzato, ha caratterizzato il secolo scorso, in presenza di skill delle professioni e dei lavori ancora catalogabili e relativamente stabili, che ora non esiste più. Ora siamo in una nuova fase: l'immateriale prevale sul materiale, l'informazionale sull'industriale; il lavoro intellettuale si ibrida con il manuale, quello autonomo con il dipendente. Il lavoro instabile e flessibile domina su tutto. Le funzioni di comando, innovazione e gestione convivono con quelle sempre più disponibili dei *knowledge workers*, o di lavoratori manuali a cui sono necessarie non irrilevanti competenze. Le conoscenze avanzate decadono rapidamente, le competenze avanzate si producono solo in *learning by doing*. Tutto questo implica una nuova visione sulle relazioni tra università, apprendimento permanente e domanda di formazione, sul ruolo delle università come incubatori in aree a densa connettività, sui conflitti tra interessi economici (brevetti e copyright, per esempio) e sociali (*crowdsourcing*, per esempio) che attraversano il nuovo scenario.

* * *

La permeabilità al flusso, più alta o più bassa (e relativa più alle forme dei media spettacolari o alle forme della vita virtuale sul web); e il grado di creatività, l'investimento nelle attività di ri/mediazione della memoria sono i due indici che oggi dovrebbero identificare l'esistenza di una soggettività attiva delle nostre istituzioni. In futuro, è probabile che esse si ricostituiscano in nuove forme ibride tra quelle che derivano dalla tradizione e le infrastrutture del web; che esse accentuino il loro potenziale di motori della virtualizzazione; che si organizzino in forma aperta e reticolare. Ma anche nel presente, i movimenti in corso già dimostrano che lo sforzo di alcune istituzioni che già si rendono permeabili e si configurano come sorgenti attive e interattive sul web è socialmente utile: la permeabilità rende infatti accessibile la memoria culturale ai nuovi attori della rete, oggi soprattutto le nuove generazioni più istruite, ma a breve termine larga parte dell'umanità; e il loro riconfigurarsi come sorgente è sintonizzato sui processi generali del *digital heritage*, dove la trasmissione del patrimonio di oggetti, riflessioni e immaginari insiste sulla ri/mediazione piuttosto che sulla conservazione e tradizione.

Per una storia dell'informatica umanistica

Tito Orlandi

Qualche tempo fa ho tracciato un panorama storico dello sviluppo dell'Informatica Umanistica, che a scopo introduttivo qui vorrei solo riassumere, rimandando per altri particolari al mio contributo *Un ultimo bilancio dell'informatica umanistica*, in corso di stampa negli atti del congresso di Montevarchi¹. Aggiungo che a quel congresso partecipava anche Annamaria Tammaro, ed esso è stato all'origine dell'Associazione di Informatica Umanistica (www.umanisticadigitale.it), in cui oggi sono poste le speranze di sviluppo futuro anche istituzionale, in Italia, di questa disciplina.

Mi sembra dunque che lo sviluppo storico si possa utilmente dividere in quattro periodi:

¹ *E-laborare il sapere nell'era digitale. Strumenti e tecniche per la gestione, la conservazione e la valorizzazione del patrimonio culturale in ambiente digitale*, Montevarchi (Arezzo), Sala della filanda alla Ginestra, 22-23 novembre 2007.

Periodo dei precursori (fino al 1960)

Per lungo tempo ci si è addirittura meravigliati che i computer potessero trovare un qualsiasi utilizzo nell'ambito delle discipline umanistiche. È dunque naturale che in questo periodo troviamo due soli progetti da prendere in considerazione. Uno è quello volto ad automatizzare il procedimento della traduzione, ed è stato proposto da William Weavers nel 1947; l'altro quello di produrre l'indice delle opere di s. Tommaso, ed è stato proposto da Roberto Busa nel 1946, per poi essere ampliato per produrre le concordanze.

Periodo delle applicazioni pionieristiche (1960-1970)

I computer cessano di essere oggetto di meraviglia, almeno presso alcuni studiosi illuminati, e intraprendenti umanisti in maggior numero cercano di sperimentare le capacità delle macchine. Indicizzazione, ordinamento, e statistica lessicale, nelle loro varie tipologie, sia pure sotto nomi variamente fantasiosi, sono il campo d'azione principale: in Italia con Alinei e Avalle; in ambiente anglosassone con Wisbey, Ellison, Morton, e Wake; in Francia con Quemada.

Periodo degli esperimenti e delle ricerche (1970-1990)

È questo, secondo me, il periodo più vivace ed interessante della storia che stiamo seguendo. Per quanto concerne i fenomeni testuali, l'analisi lessicale si amplia all'analisi delle strutture linguistiche, con la nascita della linguistica computazionale. Ricordo Zampolli, Berni Canani, Geoffrey Sampson, Susan Hockey. Quest'ultima, insieme con l'ambiente di Oxford, ha anche il merito di aver sviluppato

software pionieristici, come il linguaggio Snobol e l'Oxford Concordance Program. L'attività in questo campo fu molto vasta.

Periodo del trionfo della tecnologia (1990 ad oggi)

Gli incredibili avanzamenti della tecnologia hanno messo fra le mani degli studiosi (in senso non figurato, se pensiamo ai computer portatili) strumenti di una potenza ignorata dai più grandi mainframe degli anni precedenti, e sono oltretutto già dotati di software più o meno incorporati, che hanno fatto dimenticare i semplici ma essenziali principi su cui il loro funzionamento si basa, ed hanno limitato il loro apprezzamento ai programmi che essi offrono, eclatanti dal punto di vista dell'utente finale, col risultato di focalizzare l'attenzione sul risultato visibile, e non sulle procedure sperimentali che dovrebbero essere l'oggetto della ricerca scientifica anche per le discipline umanistiche.

Detto questo, e sottolineato il fatto che le acquisizioni più importanti sono state fatte nel periodo 1970-1990, e sono poi state sostanzialmente trascurate o peggio dimenticate, quello che mi sembra utile nel contesto di oggi, cioè del ricordo delle idee e dell'opera di Giuseppe Gigliozzi, è ricordare l'attività del cosiddetto "gruppo di Roma" all'incirca fra 1980 e 1990. Essa si svolgeva in modo particolare intorno al gruppo di ricerca "Informatica e discipline umanistiche" dell'università degli studi di Roma "La Sapienza" (Facoltà di Lettere), da me diretto. I partecipanti erano: Giovanni Adamo, ricercatore, Lessico Intellettuale Europeo C.N.R.; Marina Beer, ricercatore, Dipartimento di Italianistica; Patrizia Bertini Malgarini, Dipartimento di Italianistica; Glauco Bertocchi, Centro di Calcolo Interfacoltà; Francesco Bianchi, Centro di Calcolo

Interfacoltà; Paola Bono, Istituto di Lingua e letteratura inglese (Facoltà di Magistero); Liana Cellerino, Dipartimento di Italianistica; Angelo Cicchetti, Dipartimento di Italianistica; Giovannella Desideri, Dipartimento di Italianistica; Ivan Di Stefano Manzella, Dipartimento di Scienze dell'antichità; Riccardo Duranti, Istituto di Lingua e letteratura inglese (Facoltà di Magistero); Giuseppe Gigliozzi, Dipartimento di Italianistica; Giuliano Macchi, Dipartimento di Studi romanzi; Antonio Magi Spinetti, Centro di Calcolo Interfacoltà; Raul Mordenti, Dipartimento di Italianistica; Giovanni Ragone, Dipartimento di Italianistica; Maria Serena Sapegno, Dipartimento di Italianistica; Paolo Sensini, Facoltà di Ingegneria; Maria Vittoria Tessitore, Istituto di Lingua e letteratura inglese (Facoltà di Magistero); Anna Maria Valente Bacci, Dipartimento di Letterature comparate; Marco Veneziani, Lessico Intellettuale Europeo C.N.R.

Finalità della ricerca era l'elaborazione di una comune base metodologica e funzionale tra discipline umanistiche e informatica. I collaboratori a questo progetto provenivano da esperienze diverse di applicazione del calcolatore, ciascuno per proprio conto, e avevano

in comune la convinzione che fosse opportuno e possibile unificare le metodologie applicative e approfondirle quindi in varie direzioni.

È anche utile ricordare la situazione di quegli anni per quanto riguarda le infrastrutture.

Dei calcolatori personali erano apparsi: Apple (1976, quasi sconosciuto), Commodore (1977), IBM PC (1981) Olivetti M20 (1982), Sinclair ZX Spectrum (1982). Essi erano assai poco utilizzabili, e del resto poco usati nell'ambito della ricerca. Infatti per applicazioni

serie si ricorreva quasi esclusivamente ai cosiddetti mainframe (Univac 1100, IBM 3090, ecc.) che erano appannaggio dei grossi centri di calcolo di università e grandi aziende. Fra i due estremi si collocavano i cosiddetti *mini*, usati per esperimenti raffinati dai grandi centri di ricerca americani. Il più celebre fu il Vax PDP11, su cui fu implementato il prototipo di Unix. I centri di informatica umanistica erano molto rari (cfr. sotto). A Roma l'esperimento del Servizio per l'Informatica della Facoltà di Lettere e poi la fondazione del CISADU (Centro Interdipartimentale di Servizi per l'Automazione nelle Discipline Umanistiche²) rappresentavano lo sforzo di rendersi autonomi dai centri di calcolo centrali dell'Università, che ovviamente privilegiavano gli utenti di carattere scientifico.

Riportiamo dunque le relazioni che membri dell'IDU hanno stilato, in seguito a visite personali effettuate in quelli che apparivano allora i principali centri di ricerca di informatica umanistica. Come si vedrà, esse sono compilate secondo uno schema prefissato.

Relazione sul "Bollettino di italianistica". A cura di Graziella Pulce

Sede: Roma, Università degli Studi "La Sapienza". Facoltà di Lettere e Filosofia. Dipartimento di Italianistica.

Organizzazione: Direttore responsabile: Alberto Asor Rosa; Comitato direttivo: Alberto Asor Rosa, Ignazio Baldelli, Nino Borsellino, Pieter De Meijer; Redazione Armando Gnisci, Giorgio

² <http://rmcisadu.let.uniroma1.it>

Patrizi, Maria Serena Sapegno, Mirella Serri; Segretaria di redazione: Alessandra Mattioli; Sezione bibliografica a cura di: Patrizia Bertini Malgarini, Angela Carella, Gabriella Di Paola, Rosamaria Facciolo, Gabriella Macciocca, Paolo Procaccioli, Graziella Pulce, Tiziana Rinaldi, Luigi Trenti.

Compiti: Il Bollettino, al suo secondo numero, è un periodico semestrale che consta di due parti. La prima, di carattere saggistico, si articola in 4 rubriche: le rassegne sull'italianistica, le informazioni sui convegni e iniziative culturali, l'italianistica nel mondo, il notiziario delle ricerche.

La seconda parte sotto il titolo di *Bibliografia* raccoglie tutte le pubblicazioni (in rivista o in saggio) afferenti all'italianistica e le presenta suddivise in sezioni (Origini e Duecento, Dante, Trecento, Quattrocento, Cinquecento, Seicento, Settecento, Ottocento, Novecento, Lingua), per ciascuna sezione ordinate alfabeticamente. Le informazioni bibliografiche sono corredate di un soggettoario che ne permette un più rapido reperimento e una migliore utilizzazione. Per tutto questo ci si è avvalsi dell'uso di un elaboratore elettronico il cui programma prevede un data entry con input agevolato da schermate. Sul file così ottenuto si lavora con il text editor che organizza e ordina i dati. Il programma prevede diverse funzioni e dispone di 7 moduli di ricerca. Tutte le informazioni vanno a costituire una banca dati la cui effettiva utilizzazione è tuttavia per ora solo auspicata.

Programmi: Le informazioni sono inserite tramite schermate che ne selezionino i contenuti: 1) sezione di appartenenza, 2) autore, 3) titolo, 4) informazioni bibliografiche, 5) campo dell'indice dei nomi e o delle opere anonime che non compaiano già al n. 2, 6) rinvio ad

altre schede, 7) campo delle recensioni, con le relative notizie bibliografiche, 8) soggetto; 9) numero d'ordine. Il *Soggettario* si articola in 4 macro-sezioni: 1) opere generali; 2) secolo; 3) autore; 4) lingua. Ciascuna di queste macro-sezioni si suddivide in sottosezioni che rendono l'informazione via via più dettagliata.

Utenti: Studiosi di italianistica.

Pubblicazioni: «Bollettino di Italianistica. Bibliografia e informazione culturale», realizzato in collaborazione con il Centro di Calcolo Interfacoltà dell'Università degli Studi di Roma, pubblicato da E.J.Brill (Leiden).

Rapporto di Giovanni Adamo e Tito Orlandi sulla missione di contatto e di informazione compiuta a Cambridge (febbraio 1985)

Denominazione e sede: University of Cambridge, Literary and Linguistic Computing Centre (LLCC), Sidgwick Avenue, Cambridge CB3 9DA, England. Tel. 0223.356411, ext. 37.

Direzione e organizzazione: Director: Dr. Terence MOORE; Ass. Dir. of Research: Dr. John L. DAWSON. L'équipe comprende altri 3 assistenti. Il LLCC è un sub-dipartimento del Department of Other Languages nella Faculty of Modern and Medieval Languages.

Compiti: Formato nel 1963, il LLCC provvede a fornire i mezzi (e la relativa consulenza) per automatizzare i seguenti settori di ricerca: Concordanze; Indici; Edizioni critiche; Bibliografie; Sinossi di testi; Studi lessicali; Studi stilistici; Cataloghi. Macchine e Collegamenti: Il LLCC dipende quasi completamente dallo University Computing Service (UCS), che dispone di un main-frame IBM3081; di una stampante laser QMS Lasergraphix 1200; di un lettore ottico KDEM. Il LLCC dispone di 5 linee telefoniche per i collegamenti con l'UCS; di 5 terminali video; 2 stampanti tipo daisywheel (di cui una Diablo

con daisy- wheel "doppia"); 1 stampante tipo "dot-matrix"; 2 perforatori di nastri cartacei (per uso di data-entry locale a basso costo); 1 microcomputer BBC, utilizzabile anche come terminale.

Programmi: È stato seguito il metodo di mettere a disposizione degli utenti dei packages di programmi standard, che essi possono usare singolarmente. Essi sono:

- ZED, Text-Editor (prodotto dall'UCS)
- CAMTEXT, Concordanze di vario tipo (prodotto dal LLCC)
- OCP (Oxford Concordance Program; vedi oltre la scheda di Oxford)
- MUSCAT, Data Base soprattutto per materiali di Musei. Ne esiste un estratto per piccole quantità di dati, inseribile in microcomputer, chiamato MUSCATL
- GCAL, Text-formatting
- SPSS e altri packages statistici.

Utenti: "Members of the Regent House, research students, and certain visitors, attached to the Faculties of Classics, Divinity, English, History, Law, Modern and Medieval Languages, Music, Oriental Studies, and Philosophy". Un elenco di utenti con i relativi studi in preparazione o condotti a termine è nella relazione depositata nel nostro Archivio.

Possibili collaborazioni: Scambio di informazioni.

Pubblicazioni: Non vi è un Bollettino regolare del LLCC. Vengono fatte delle Note o dei Rapporti ufficiali, che vengono policopiati. Da ricordare tuttavia i Proceedings del Meeting tenutosi in occasione del ventesimo anniversario della fondazione del LLCC, di prossima pubblicazione.

Rapporto di Giovanni Adamo e Tito Orlandi sulla missione di contatto e di informazione compiuta a Catania (Cattedra di Letteratura italiana moderna e contemporanea; CLIPON; CED della SIFI), 17-18 maggio 1984

Sulla base di contatti precedentemente intercorsi con alcuni ricercatori della Facoltà di lettere dell'università di Catania, si è venuti a conoscenza dell'attività di elaborazione di testi letterari, per lo più poetici, intrapresa dalla Cattedra di Letteratura italiana moderna e contemporanea, di cui è titolare il Prof. Giuseppe Savoca. Da questa iniziativa, con l'apporto di studiosi di altre università, ha preso l'avvio un "Gruppo Nazionale di Coordinamento" del CNR, il CLIPON. Le operazioni di calcolo vengono svolte in collaborazione con il Centro di elaborazione dati di un'industria farmacologica catanese, la SIFI.

1. Cattedra di Letteratura italiana moderna e contemporanea dell'Università degli studi di Catania.

Sede: Istituto di Letteratura Italiana, (Palazzo Sangiuliano), Piazza Università, 95124 Catania. Tel. 095/326757.

Direzione e Organizzazione: Direttore della Ricerca è il titolare della cattedra di Letteratura italiana moderna e contemporanea, prof. Giuseppe Savoca. Direttore dell'Istituto è il Prof. Carmelo Musumarra. Collaborano alla ricerca i dott. Sebastiano Grasso e Salvatore Baglio.

L'analista, che ha provveduto alla stesura dei programmi, il dott. Giovanni Pandolfo, è il direttore di un Centro di elaborazione dati esterno (vedi, CED della SIFI). Egli non aveva in precedenza alcuna esperienza di applicazioni dell'informatica alle discipline umanistiche. Si prevede per il futuro l'acquisizione di competenze

specifiche nel settore della programmazione da parte del personale della ricerca.

Attività: L'inizio dei lavori risale al 1977, nell'ambito di esperienze di critica letteraria, nel corso delle quali si avvertiva il bisogno di disporre di strumenti più precisi. Nel corso di questi anni sono state stabilite liste di concordanza di vari testi dell'Otto e Novecento. Alcune di esse costituiscono il lavoro di studenti laureandi.

Macchine: Presso l'Istituto è installato un sistema di scrittura TES Olivetti con un piccolo programma di information retrieval per la gestione di una bibliografia di italianistica. È in arrivo il sistema 36 IBM, con cui si provvederà alla gestione in sede dei programmi di concordanza (attualmente viene usato il sistema 34 IBM del CED-SIFI).

Collegamenti: Gli studiosi dell'università di Catania possono accedere al terminale esistente presso l'Istituto di astronomia, collegato con il CINECA di Bologna. Si hanno contatti saltuari con la IBM per assolvere a particolari esigenze tecniche.

Programmi: È stato messo a punto dal Dott. Pandolfo un programma di concordanza di testi letterari, che prevede una fase di intervento manuale per l'assegnazione ad ogni forma del testo del relativo lemma e della categoria grammaticale, secondo un modello predisposto dal prof. Savoca. Si prevede di rendere presto parzialmente automatica la fase di lemmatizzazione, grazie alla creazione di "inverted files" delle forme non omografe già riscontrate in precedenti elaborazioni. Per le forme omografe si procederà al loro raggruppamento e all'assegnazione in automatico del lemma. Le forme non ancora attestate negli spogli continueranno a ricevere

manualmente il lemma e la categoria grammaticale. Si è inoltre dato corso a un progetto, già parzialmente attuato, di costituire una piccola banca dati di italianistica.

Possibili collaborazioni: La ricerca è aperta ad ogni tipo di possibile contatto, anche se non sono stati ancora individuati settori precisi di collaborazione.

Pubblicazioni: Nessuna riguardante la metodologia. Si può ottenere a richiesta l'elenco delle realizzazioni già attuate. Per quanto riguarda i risultati, è in corso di stampa la concordanza di: *Tutte le poesie* di Guido Gozzano, a cura di G. Savoca.

2. CLIPON (Concordanze della Lingua Italiana Poetica dell'Otto e Novecento).

Sede: Istituto di Letteratura italiana dell'Università degli Studi di Catania (cfr. sopra).

Direzione e Organizzazione: Gruppo Nazionale di coordinamento, del CNR. Università contraenti: Cagliari, Catania, Lecce, Roma "La Sapienza", Roma "Tor Vergata", Torino, Urbino. Comitato direttivo: Proff. Maxia (CA), Savoca (CT), Bonea (LE), Petrucciani (RM I), Scrivano (RM II), Guglielminetti (TO), Cerboni Baiardi (Urbino). Esperti: Proff. Ignazio Baldelli (RM I), Roberto Busa, Alfredo Stussi (PI), Alvaro Valentini (MC), Antonio Zampolli (PI).

Compiti: Il Gruppo, costituito nel corso di un incontro tenutosi a Urbino nel settembre del 1983, si propone di coordinare le attività svolte in campo nazionale nel settore specifico delle concordanze della lingua italiana poetica. Il fine ultimo è quello di costituire un lessico della poesia italiana contemporanea. L'indirizzo attuale, per quanto riguarda l'organizzazione pratica del lavoro, è quello di provvedere alla registrazione dei dati presso i Centri menzionati

sopra, e di produrre le concordanze a Catania. È previsto in futuro un collegamento via cavo di terminali dai vari Centri con Catania.

Pubblicazioni: Sono allo studio le modalità di diffusione e pubblicazione delle concordanze ottenute e la possibilità di pubblicare una rivista del Gruppo (si pensa all'eventuale titolo di «Lemma»).

3. Centro di Elaborazione Dati (CED) della SIFI.

Sede: Via Nicola Coviello 15/b, 95128 CATANIA. Tel. 095/438178.

Direzione e organizzazione: Direttore, Dott. Giovanni Pandolfo. Centro di servizio di attività commerciale privata.

Compiti: Elaborazione e gestione dati per la SIFI e per utenti esterni (Enti e privati). Noleggio terminali.

Macchine: Sistema 34 IBM a cui sono collegati anche terminali di utenti esterni.

Programmi: Linguaggio di scrittura: RPG. Presso il CED si provvede attualmente anche alla registrazione dei dati per ottenere le liste di concordanza.

4. Centro di calcolo dell'Università degli Studi di Catania.

Accanto al Centro di calcolo già esistente, destinato a compiti amministrativi, è in fase di organizzazione un Centro di servizi interdipartimentale, che verrà dotato di un sistema UNIVAC 1100.

Rapporto di Tito Orlandi sulla missione di contatto e di informazione compiuta a Louvain-la-Neuve, presso il CETEDOC (giugno 1984)

Denominazione e sede: CETEDOC = Centre de Traitement Electronique des Documents. Université Catholique de Louvain-la-Neuve, Collège Erasme, Place Blaise Pascal 1,B - 1348 Louvain-la-Neuve. Tel. 010.418181 int. 3788.

Direzione e organizzazione: Direttore prof. Paul Tombeur (tiene corsi di: latino medievale; paleografia latina; informatica). Laboratorio di ricerche metodologiche, insegnamento e servizi nell'ambito dell'informatica applicata alle discipline umanistiche. Dipende direttamente dal Doyen della Faculté de lettres e dal Conseil académique.

Compiti: Fondato nel 1968 per assistere studenti e ricercatori in vista dell'utilizzazione dell'ordinatore nel campo delle discipline umanistiche, particolarmente per tutto ciò che concerne lo studio dei testi o documenti similari, quali gli archivi o la documentazione relativi all'archeologia e storia dell'arte. I metodi messi a punto si concretizzano nell'elaborazione di programmi (cfr. sotto). L'informatica ed i metodi propri a ciascun tipo di ricerca sono tema di corsi e di seminari. I programmi sono messi a disposizione di tutti coloro che lavorano in collaborazione con il CETEDOC, dando origine a progetti in comune, eventualmente finanziati dal Fonds national de la recherche scientifique. Parecchi lavori sono anche realizzati in collaborazione internazionale.

Macchine: Una macchina IBM serie 1, di 512 K; due unità a nastro; unità a disco (rigido); unità a dischetti; stampante a punti; molti terminali. I terminali sono predisposti in modo da poter riprodurre qualsiasi tipo di carattere, ed anche la notazione musicale.

Lo stesso vale per la stampante a punti. Per il sistema COM (Computer Output in Microfiche) si ricorre ad un centro di servizi esterno. La stessa cosa vale per la fotocomposizione. Ambedue vengono attuati a partire da nastri prodotti presso il CETEDOC.

Collegamenti: Con il Centre de calcul dell'università (macchine IBM) tramite CMS (Conversational Monitoring System). I programmi più impegnativi sono fatti girare al Centre de calcul, mentre le macchine del CETEDOC sono usate soprattutto in fase di input, di correzione dei files, e per programmi di minore impegno. Presso il Centre de calcul opera anche una stampante laser, usata per i lavori che non richiedono la fotocomposizione. Anch'essa può produrre caratteri di vario tipo, generati da programmi che vengono studiati nell'ambito del CETEDOC. Per i collegamenti con centri più lontani (in Belgio o all'estero; specialmente la Francia) si preferisce al momento ricorrere allo scambio di nastri, in quanto i collegamenti on-line non danno risultati soddisfacenti in proporzione alla spesa.

Programmi: Il CETEDOC usa esclusivamente programmi prodotti al suo stesso interno, dunque non packages e simili. Questo vale a partire dal programma di editing, fino alla creazione dei files, e al trattamento documentario e linguistico. Il trattamento linguistico riguarda soprattutto le concordanze. Esse sono condotte in modi vari, tramite un programma costituito di parti differenti, che consentono trattamenti a partire dalla semplice lista delle forme, via via verso modi più sofisticati: calcoli statistici sulle forme; parole nel contesto; variazioni nella scelta e nella presentazione del contesto; ecc. La lemmatizzazione è fatta in modo parzialmente automatico, tramite dizionari di macchina delle varie lingue (soprattutto il latino) che si stanno costituendo. Altro tipo di programmi a cui è stata

dedicata particolare cura è quello relativo alla visualizzazione e alla stampa dei caratteri non latini (soprattutto greco e arabo). Tuttavia, per quanto riguarda la codifica dei testi che devono essere sottoposti a tali programmi, non sono disponibili manuali stampati. Non è previsto lo scambio di programmi con altri centri interessati.

Utenti: Il Centro è rivolto soprattutto agli studenti e ricercatori dell'Université catholique di Louvain-la-Neuve. Ma è anche fortemente orientato verso la collaborazione con tutte le imprese in campo internazionale, soprattutto quelle che si occupano di trattamento dei testi. Tale collaborazione si esplica in sostanza nella memorizzazione di archivi di testi (che possono essere scambiati etc.), il cui trattamento deve essere però condotto al CETEDOC (per quanto gli compete, coi suoi propri programmi). Il lavoro si viene così a costituire in "progetti", con varie finalità e organizzazione. I più importanti sono:

"Tradition occidentale latine", diviso in sotto-progetti.

- Thesaurus patrum latinorum. Trattamento sistematico degli scritti dei Padri della Chiesa latina e di opere del Medio Evo latino. È condotto nel quadro del Corpus Christianorum dell'editore Brepols (cfr. sotto).
- Dictionnaire du latin médiéval belge, nel quadro del Novum Glossarium sotto il patronato dell'Union Académique Internationale.
- I concilii ecumenici, in collaborazione con il Centre d'histoire médiévale di Paris-Sorbonne.
- Le fonti francescane e domenicane.
- Opere filosofiche latine (Avicenna latinus, Spinoza, ecc.).

Greco. Opere etiche di Aristotele; Pseudepigrifi greci dell'Antico Testamento; Storici greci di Roma; opere di Gregorio di Nazianzo.

Arabo. Avicenna.

Lingue europee. Portoghese: analisi dei testi di Camoens (tutti già memorizzati).

Olandese: romanzi di Hella Haasse. Progetti minori in altre lingue.

Musica.

Possibili collaborazioni: Si veda quanto detto sopra. Dal punto di vista dell'IDU sarà particolarmente importante lo scambio di notizie e teorie sulla metodologia del trattamento dei testi e della visualizzazione su schermo e a stampa.

Pubblicazioni: È disponibile, per uso interno, l'indice per autore dell'archivio di testi memorizzati presso il CETEDOC. Descrizione analitica del CETEDOC nell'articolo *Bilan des travaux et des recherches du CETEDOC*, che sarà pubblicato in «Informatique et Sciences Humaines». Altre pubblicazioni sul tema sono disponibili in archivio presso l'IDU.

Rapporto di Tito ORLANDI sulla missione di contatto e di informazione compiuta a Liège, CIPL (ex LASLA). 22 giugno 1984

Denominazione e Sede: CIPL = Centre Informatique de Philosophie et Lettres. Université de Liège, Faculté de Philosophie et Lettres (Résidence A. Dumont, 7^e étage). Place du Vingt Aou[^]t, 32, 4000 Liège. Tel. 041.420080.

Direzione e organizzazione: Direttore Prof. Joseph Denooz. Conseil de Gestion: Presidente. prof. Cl. Desama ; membri: Proff. Hélin, Evrard, Rutten, Lavis, Noe"l, Limet, Gérin, Otte. Rappresentanti del personale scientifico e del personale amministrativo; delegato del S.E.G.I. (Service Général d'Informa-tique); segretario. Personale scientifico: S. Govaerts. Personale amministrativo e tecnico: M. W. Stassen (programmatore); M. C. Lochen (operatrice); C. Delmotte (segretaria).

Compiti: Creato per decisione del Consiglio d'amministrazione dell'università in data 26 ottobre 1983 (come trasformazione del LASLA = Laboratoire d'Analyse Statistique des Langues Anciennes, operante presso la medesima Università circa dal 1960), il CIPL è una unità decentrata del SEGI (cioè del Centro di calcolo centrale dell'università). Il compito assegnato al CIPL è quello di aiutare i servizi della Facoltà di Filosofia e lettere:

- fornendo il materiale informatico meglio adatto alle loro necessità
- dando consigli per lo sviluppo e la messa a punto di programmi od offrendo loro programmi già realizzati
- intervenendo presso il SEGI per questioni di finanziamento.

Il CIPL cura anche l'informazione e la formazione degli utilizzatori e degli studenti attraverso tre tipi di attività:

- Corso di BASIC (45 ore)
- Corso per il Certificat d'Etudes en analyse automatique de l'Information en Sciences humaines (210 ore di corsi)
- Ciclo di formazione per l'insegnamento assistito dal calcolatore (30 ore), destinato specialmente agli insegnanti di scuole secondarie.

Macchine:

- Macchine connesse al calcolatore IBM 370/158 del SEGI: 5 terminali IBM 3276 e IBM 3278. 1 telescrivente IBM 2741 adattata al trattamento della lingua greca, adatta a stampare documenti di alta qualità per l'off- set. 2 stampanti IBM 3289 (300 linee al minuto). N.B.: Tutte le modifiche di tastiera e configurazione video dei terminali (che sono in affitto) sono state studiate dalla IBM belga e messe in opera dalla IBM irlandese.
- Macchine autonome: Personal SORD (2 disk-drivers; stampante ad alta qualità); 3 Personal portatili OSBORNE 1 (2 disk-drivers; stampante); 1 Personal portatile OSBORNE Executive (2 disk-drivers; stampante). Questo apparecchio è dotato di EPROM per lo schermo, per cui si possono definire vari caratteri a volontà dell'utente.

N.B.: Le macchine autonome sono state messe in grado di comunicare col mainframe del SEGI, ma l'organizzazione deve essere perfezionata (finora si agisce con l'accoppiatore acustico, coi noti problemi di lentezza e scarsa affidabilità).

Collegamenti: Avvengono tramite il SEGI. Attualmente esiste una rete generale per la città di Liegi. Si sta mettendo a punto il

collegamento col CIRCSE (regione parigina, CNRS). Ma è considerato acuto il problema dei costi.

Programmi: I programmi che girano sul main-frame del SEGI sono prodotti direttamente dal personale del CIPL (e prima del LASLA), con la sola eccezione dell'editor (per scrivere i files), per il quale si usa lo standard IBM (procedure TSO e STP). In particolare si menzionano i programmi per:

- lemmatizzazione ed analisi automatica per varie lingue
- rilevazioni statistiche di dati
- stampa di indici e concordanze
- costituzione di repertori bibliografici
- analisi automatica di professioni e provenienze; analisi di liste nominative di censimento (cfr. sotto, utenza di storici).

Sui micro-calcolatori girano programmi di word- processing e di insegnamento assistito (latino; francese).

Utenti: Il CIPL si rivolge soprattutto al personale scientifico e agli studenti della Facoltà di Lettere e Filosofia. Per beneficiare dei suoi servizi occorre ottenere un "crédit-calcul", prendere in carico le spese specifiche del lavoro (p.es. carta etc.), e, nella misura del possibile, provvedere in proprio all'inserimento dei testi e dei dati e alla correzione dei file. Per prestazioni speciali il CIPL richiede uno speciale contributo finanziario.

Fra le imprese si menzionano:

- Prof. Ruten, indici di Aristotele (Metafisica, Fisica, etc.).
- Prof. Halleux, indici e concordanze parziali degli Alchimisti greci.
- Equipe del P. Costabel, studi su Decartes (storia delle scienze).

- "Theorema": progetto di studio di ricette mediche antiche in varie lingue, con interconnessione tramite thesauri (formazione di un data-base).
- Prof. Desama, studi di modelli demografici e di storia economica (Belgio, XVII-XIX sec.).
- Memorizzazione di registri parrocchiali, notarili ecc. da parte di un assistente dell'Archivio generale belga.

Possibili collaborazioni: Il Centro è orientato verso lo scambio di file di testi. È invece contrario allo scambio di programmi. È favorevole allo studio di standard internazionali per memorizzare testi. Però c'è scetticismo a causa della dispersione degli studiosi che si avvalgono dell'informatica. D'altra parte dovrebbe essere sempre possibile operare delle transcodifiche "a posteriori".

Pubblicazioni: Brossura descrittiva del CIPL (depositata nell'archivio IDU).

Indici delle parole di: Virgilio, Seneca, Corpus Hermeticum, Cicerone Phil., Tacito.

Concordanza di Vitruvio.

Lexique inverse de Plaute.

Lexique de base du latin; Dictionnaire fréquentiel et index inverse de la langue latine. L'ordinateur et le latin; Actes du Congrès *Informatique et Sciences humaines*.

Rivista: «Informatique et Sciences Humaines».

Scheda informativa relativa a: Opera del Vocabolario dell'Accademia della Crusca

Sede: c/o Accademia della Crusca Via di Castello, 46, 50141 Firenze

Direzione e Organizzazione: L'Opera del Vocabolario è diventata da alcuni anni Centro di Studio del CNR. Precedentemente alla trasformazione in Centro CNR, la direzione dell'Ufficio Lessicografico dell'Accademia era affidata al Prof. Silvio D'Arco Avalle.

Compiti: In vista della redazione di un Vocabolario storico della lingua italiana, si è previsto di registrare in memoria un corpus di testi letterari italiani a stampa, per ognuno dei quali viene stabilita una lista di concordanze da sottoporre a lemmatizzazione. Per l'elenco dei testi spogliati, cfr. Pubblicazioni.

Macchine: È prevista l'istallazione di un terminale collegato con il CNUCE di Pisa, presso il quale si svolgono attualmente tutte le elaborazioni.

Programmi: I programmi di trattamento dei testi sono stati compilati in collaborazione con il personale dell'Istituto di linguistica computazionale del CNR (Pisa), da un programmatore dell'Opera del Vocabolario distaccato a Pisa.

Pubblicazioni:

- D'Arco Silvio Avalle, *Al servizio del Vocabolario della Lingua Italiana*, Firenze, Accademia della Crusca, 1979
- Lista delle opere sottoposte a spoglio.

Rapporto di Paola Bono sulla missione informativa e di contatto compiuta a Edinburgh (Scotland), Settembre 1984

1. Department of computer science, University of Edinburgh.

Sede: James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, Gran Bretagna.

Direzione e organizzazione: Direttore del Dipartimento (Head of Department) è attualmente il Dott. Peter Schofield, ma il Dipartimento è una struttura didattica e di ricerca assai ampia con varie articolazioni, ognuna delle quali fa capo a un diverso coordinatore.

Compiti: Creato nel 1966 in seguito alla ristrutturazione della Computer Unit, ha responsabilità didattiche e di ricerca nel campo della Computer Science. Nel 1966 venne istituito anche l'ERCC (Edinburgh Regional Computing Centre) come struttura di servizio per l'università e la ricerca nell'area di Edinburgh. Da allora l'autonomia operativa del Dipartimento è cresciuta man mano che si espandeva la sua dotazione di macchine e di personale.

Macchine: Il servizio di computer l'università è fornito dall'ERCC, che dispone a tale scopo di un sistema gemellare ICL 2972, collocato nella sede del Department of Computer Science. Da questa installazione dipende l'EMAS (Edinburgh Multi-Access System) che può servire oltre 100 utenti simultaneamente impegnati su terminali interattivi situati in diversi edifici dell'Università e collegati da una rete locale. La rete dà accesso a vari altri elaboratori, incluso un altro ICL 2900, e a molte stampanti, compositrici, ecc. Per specifici progetti di ricerca si può utilizzare il DEC KL-10 dello Science Research Council; per altri vi sono macchinari a loro esclusivamente destinati. Il Dipartimento sta organizzandosi per

condurre la maggior parte del suo lavoro su elaboratori personali progettati nel Dipartimento stesso, con un "filestore" comune su una rete di comunicazione tipo Ethernet. Al momento ce ne sono 40, basati su un elaboratore M6 8000 con 1/2 megabyte di memoria. C'è poi un VAX 11/780 installato in sede, con 4 megabyte di main store e 1200 megabyte di disk storage. Serve circa 40 utenti in simultanea su VMS ed è collegato all'ERCC. Linguaggi PASCAL, IMP, FORTRAN. Il Dipartimento dispone inoltre di vari minicomputer, alcuni intesi per compiti specializzati.

Programmi: Elaborati all'interno del Dipartimento.

Utenti: Docenti, personale e studenti del Dipartimento. Sono in atto programmi di collaborazione con istituzioni culturali pubbliche e private; nel campo umanistico (a) con biblioteche per l'impostazione e la gestione di sistemi computerizzati di catalogazione e classificazione; (b) con musei e istituti di archeologia per elaborare sistemi computerizzati di classificazione dei reperti.

Possibili collaborazioni: All'interno del Dipartimento si interessano di problemi legati al campo disciplinare umanistico, e in particolare di analisi stilistica e determinazione autoriale: il prof. Sidney Michaelson (prof. of Computer Science), il Dr. Andrew Morton (membro della Research Staff) e Thomas Horton (post-graduate student per un Ph.D.). Per iniziativa di T. Horton si è formato all'inizio del 1984 un Computer in the Humanities Club, che include docenti e studenti di vari dipartimenti (Storia, Letteratura Inglese, ecc.), alcuni già con una certa esperienza e conoscenza del campo, altri interessati ad acquisirla. T. Horton ha assicurato che ci terrà informati sulle attività del club, fornendoci innanzi tutto una lista dei membri e dei loro interessi di ricerca.

2. Department of artificial intelligence, University of Edinburgh.

Sede: Forrest Hill, Edinburgh EH1 2QL, Scotland.

Direzione e organizzazione: Direttore (Head of Department) è attualmente il Dott. J. A. M. Howe.

Compiti: Fondato nel 1966, è l'unico dipartimento di intelligenza artificiale in Gran Bretagna. Ha compiti didattici e di ricerca.

Macchine: Si serve in parte del DEC KL-10 dello Science Research Council e del sistema EMAS dell'ERCC. Dispone inoltre di un sistema VAX 11/750 Berkeley 4.1 BSD UNIX, e di un sistema GEC serie 63 che utilizza l'UX63.

Programmi: Elaborati nel Dipartimento stesso, che elabora anche linguaggi specifici per la ricerca nel campo dell'intelligenza artificiale (per esempio, POP2 e PROLOG).

Utenti: Docenti, personale e studenti del Dipartimento. Sono in atto progetti di collaborazione con scuole di vario ordine e grado e con complessi industriali, per mettere a punto programmi, linguaggi e robot appositamente elaborati e costruiti.

Pubblicazioni: Sono già state pubblicate più di 200 relazioni e comunicazioni che possono essere acquistate presso il Dipartimento; cfr. Allegati in archivio per gli abstracts di quasi tutti quelli pubblicati tra il 1979 e il 1984.

3. School of epistemics, University of Edinburgh.

Sede: 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland.

Direzione e organizzazione: Direttore (Director) è attualmente il Dott. E. B. Richards.

Compiti: Fondata nel 19868 come unità interdisciplinare tesa ad esplorare l'area della scienza cognitiva. Struttura di ricerca, offre anche un programma di studi post-graduate per il conseguimento di Master o Ph.D. Include docenti dei dipartimenti di filosofia, psicologia, linguistica e intelligenza artificiale. Il computer è utilizzato nella ricerca di modelli cognitivi.

Macchine: Un VAX 11/750, tre VAX/750, diversi microprocessori.

Utenti: Docenti, personale e studenti della School.

4. Department of Linguistics, University of Edinburgh.

Sede: Adam Ferguson Building, George Square, Edinburgh EH3, Scotland.

Macchine: Un VAX 11/750 con sistema operativo VMS.

Nota: L'informazione su questo Dipartimento è incompleta in quanto non ho avuto tempo sufficiente per prendere direttamente contatto con qualche membro del Dipartimento e per procurarmi il materiale informativo ad esso relativo. Nel Dipartimento di Linguistics si fa uso dell'elaboratore per la ricerca sulle lingue naturali e per la costruzione di modelli formali. Anche il Department of English literature dispone di diversi micro per word-processing.

Rapporto di Tito ORLANDI sulla missione di contatto e di informazione compiuta a Milano presso il padre Prof. Roberto BUSA, 28-29 dicembre 1984

1. Index Thomisticus

Sede: Fino al 1985 non aveva sostanzialmente una sede propria, ma era in certo modo

legato alla residenza del Padre Busa, fondatore e direttore (cfr. sotto). Dal 1985 la sede

"legale" può essere considerata quella delle Facultés universitaires "Notre Dame de la

Paix" a Namur (fondazione universitaria dell'ordine dei Gesuiti).

Direzione e organizzazione: I direttori sono attualmente il Padre Roberto BUSA (c/o Coll. Aloysianum, Gallarate) e il Padre Jacques Berleur, ora rettore a Namur (cfr. sopra), studioso di informatica.

L'impresa comprende due settori:

- Una società che gestisce gli utili provenienti dalla vendita dei 56 volumi dell'Index Thomisticus.
- Un'associazione, CAEL (Associazione per la computerizzazione delle analisi ermeneutiche lessicologiche), presieduta dal dott. Antonio Cacciavillari.

Compiti (e notizie storiche): Il progetto è nato nel 1946, ed ha iniziato l'elaborazione su macchine IBM nel 1949 (prima di tipo elettro-meccanico, poi elettronico). Ha concluso la prima fase della sua attività con la pubblicazione dei 56 volumi dell'Index Thomisticus. Nel frattempo ha sviluppato altre attività collaterali, all'inizio con lo scopo di verificare alcune metodologie applicate all'I.Th. su altro genere di testi, in particolare con diversi alfabeti:

ebraico, cirillico, greco. Ora i compiti sono di sfruttamento dei testi fin qui memorizzati (oltre a S. Tommaso, testi ebraici di Qumran, Index Aristotelicus del Bonitz, Canti della Divina Commedia, Goethe Farbenlehre vol. III, Prolegomena di Kant, Vulgata Latina), e ricerche teoriche sulle applicazioni del computer nell'ambito della linguistica. Per questo vengono organizzate e finanziate borse di studio dal CAEL; per esempio, analisi lessicologica del lemmario dell'Index Thomisticus - Segmentazione monotematica del lemmario (Raccolta automatica delle correlazioni grammaticali elementari dirette).

Macchine: Il lavoro è stato sempre svolto presso centri "privati" di calcolo, provvisti di macchine IBM di vario tipo. Ora si appoggia anche all'organizzazione del GIRCSE (vedi scheda apposita).

Programmi: I programmi sono stati sviluppati autonomamente dal gruppo di ricerca, e sono suoi particolari. Le ricerche previste nel recente piano di sviluppo vengono condotte

normalmente a partire da schede cartacee (prodotte da stampante) e poi le annotazioni finali fatte su di esse verranno aggiunte al lemmario memorizzato.

Possibili collaborazioni: L'archivio testi memorizzati consiste di circa 35 nastri a densità 6250 (per l'elenco cfr. sopra). I testi sono disponibili a prezzo di costo per ogni gruppo di ricerca che si impegni a non usarli commercialmente.

Contatti informativi: Scambio di file di testi.

Pubblicazioni: Index Thomisticus, 56 voll.

2. G.I.R.C.S.E. Gruppo Interdisciplinare Ricerche per la Computerizzazione dei Segni dell'espressione

Sede: Milano, Università Cattolica del Sacro Cuore.

Direzione e organizzazione: Presidente: Prof. Corrado Bolognesi, direttore dell'Istituto di Glottologia dell'Università Cattolica; Consiglio direttivo: Direttore, P. Roberto Busa; membri: 1 rappresentante del CILEA (vedi scheda apposita); direttore del centro di calcolo dell'Università Cattolica (Busto Arsizio); Prof. Cupis; Prof. Lucia Vian (segretaria).

Compiti: Il Gruppo di ricerca è nato nel 1978 come seminario per ricercatori presso l'Università Cattolica (dal 1974 circa esisteva un Seminario di Lessicologia e lessicografia presso l'Aloysianum di Gallarate; e presso l'Università Gregoriana di Roma). Si occupa dell'applicazione del computer agli studi di lessicografia e linguistica generale. Dal 1983/1984 è istituito un corso di insegnamento per studenti, denominato di Linguistica matematica e computazionale (sul modello di quello di Zampolli a Pisa). Conta circa 50 iscritti; e sono state chieste 25 tesi di laurea. Sono state istituite collaborazioni con: Università di Algeri; Università di Helsinki (lavoro su Folk-Songs argentini); Compagnia del Gesù (scritti di sant'Ignazio). Preparazione di concordanze contrastive multilingui.

Macchine: Il Gruppo di ricerca si avvale di terminali presso l'Università Cattolica, collegati sia al CILEA (vedi scheda apposita; macchine SPERRY UNIVAC), sia al Centro di calcolo dell'Università Cattolica, situato a Busto Arsizio (macchine IBM; la sig. Saporiti segue i lavori umanistici). Per l'input viene usato uno speciale microcomputer messo a punto dalla HYPHEN di Bologna, denominato PAMPHILUM. È munito di EPROM intercambiabile per

la gestione del video, e può quindi mostrare sullo schermo molti tipi di alfabeto (in particolare finora latino, arabo, ebraico, greco, armeno. Si basa su 124 monocodici (bytes) e 90 bicodici per l'identificazione dei segni. È allo studio l'estensione a 220 del numero dei monocodici. Vengono usate tastiere intercambiabili. È allo studio l'uso di una tastiera unica con descrizione intercambiabile dei segni alfabetici.

Programmi: Vengono scritti caso per caso.

Possibili collaborazioni: Scambio di informazioni.

Pubblicazioni: Sono in preparazione le dispense del corso, che costituiranno una introduzione alla linguistica computazionale e in generale all'informatica per le discipline umanistiche.

Scheda informativa relativa a: Lessico intellettuale europeo - Centro di Studio del C.N.R.

Sede: c/o Istituto di Filosofia, Villa Mirafiori - Via Nomentana, 118, 00161 Roma - Tel.: (06) 855.350

Direzione e organizzazione: Il Lessico Intellettuale Europeo (L.I.E.) è sorto nel 1964 come Gruppo di Ricerca e si è trasformato nel 1970 in Centro di Studio del Consiglio Nazionale delle Ricerche. Si avvale di un'équipe di ricercatori del C.N.R. e della collaborazione di docenti messi a disposizione dall'Università di Roma. È diretto dal prof. Tullio Gregory, il quale divide la responsabilità dei programmi con un Consiglio scientifico presieduto dal prof. Eugenio Garin e composto dai proff. Ignazio Baldelli, Paul Dibon (dell'E.P.H.E. di Parigi), Aldo Duro, Marta Fattori, Giovanni Pugliese Carratelli, Valerio Verra e dal Dott. Pietro Pimpinella.

Compiti: Le attività di ricerca interessano prevalentemente l'ambito della storia delle idee e del pensiero filosofico e scientifico e si propongono di fornire strumenti e contributi per l'analisi e la ricostruzione delle terminologie peculiari di studiosi, filosofi e scienziati, nelle differenti aree culturali e linguistiche europee.

Programmi particolari:

a. *Lessico filosofico dei secoli XVII e XVIII.*

Per la redazione di quest'opera, basata sullo spoglio lessicografico di oltre trecento testi filosofici appartenenti al periodo storico considerato e ripartiti in sei ambiti linguistici diversi (latino, inglese, francese, tedesco, spagnolo e italiano), si è costituita una banca di dati in cui far confluire il fondo di documentazione raccolto. Per l'ambito linguistico sono attestate oltre 8.000 entrate lessicali, per un complesso di 1.800.000 occorrenze di parole.

b. *Lessici, indici, concordanze.*

Si tratta di strumenti lessicografici ottenuti sulla base di spogli lessicali automatici di testi. L'uso del calcolatore consente infatti di estendere a livelli diversi l'analisi lessicografica di un testo, e di costituire strumenti essenziali per una diversa lettura del testo e per un attento esame dei dati in esso contenuti. Le procedure di spoglio seguono il modello di trattamento automatico di testi in lingua naturale predisposto e operante presso l'Istituto di Linguistica computazionale del C.N.R. di Pisa. Si sta adottando sperimentalmente il programma di concordanze e indici "Glossa", operativo presso il Centro di calcolo interfacoltà dell'Università di Roma "La Sapienza".

c. Thesaurus Mediae et Recentioris Latinitatis.

Il progetto di costituire un Thesaurus Mediae et Recentioris Latinitatis è stato presentato dal L.I.E. nel corso del II Colloquio Internazionale (7-9 gennaio 1977). Il contenuto della proposta permette in un primo tempo di mettere insieme i risultati delle ricerche che si vanno realizzando in vari Centri per costituire un lemmario della terminologia che può considerarsi specificamente medievale e moderna, in quanto non registrata nei lessici del latino antico e tardo- antico. Per costituire tale lemmario aperto, organizzato come banca di dati lessicali, è stata adottata una procedura informatica di tipo interattivo su "mass storage", predisposta grazie all'apporto tecnico dell'I.L.C. di Pisa.

Macchine:

- c/o Centro di Calcolo Interfacoltà dell'Università di Roma "La Sapienza": UNIVAC 1100/82, Operating System: Exec/1100 lev. 37R2
- c/o CNUCE - Via S. Maria, 36 - 56100 PISA: IBM 3033 e 370/168
- c/o L.I.E.: Video-terminali UNIVAC UTS20 e UTS40 (64K) (L'UTS40 è utilizzabile sia in locale, sia collegato con il "main frame" del Centro di Calcolo Interfacoltà). Unità floppy-disk collegata con UTS40. Local Printer UNIVAC 0798 Printer Subsystem.

Programmi:

- In ambiente UNIVAC: "Glossa" (software per spogli lessicali). "UNIDAS" ("storage and information retrieval system"). Software di interfaccia ai packages commerciali sviluppato presso il L.I.E. Sono in allestimento programmi per analisi di tipo quantitativo.

- In ambiente IBM: Software batch per l'analisi lessicale di testi in linguaggio naturale e procedure interattive per la gestione del Thesaurus Mediae et Recentioris Latinitatis, predisposti dall'Istituto di Linguistica Computazionale del C.N.R. di Pisa.

Utenti: Studiosi e ricercatori che ne facciano richiesta.

Collaborazioni: Istituto di Linguistica computazionale del C.N.R.

- Pisa; Accademia della Crusca - Firenze; Centro di studio per l'Opera del Vocabolario - Firenze; Centre de Traitement Electronique des Documents (CETEDOC) - Université Catholique - Louvain-la-Neuve (Belgio); Centre d'informatique de philosophie et lettres, Université de Liege (ex LASLA); Trésor de la langue française (T.L.F.) – Nancy (Francia); Equipe Descartes - Paris (Francia); Centre d'histoire des sciences et des doctrines del CNRS - Paris (Francia).

Pubblicazioni:

1. Giuseppe Sermoneta, *Un glossario filosofico ebraico-italiano del XIII secolo*, 1969.
2. Giovanni Crapulli, *Mathesis universalis. Genesi di un'idea nel XVI secolo*, 1969.
3. G. Crapulli - E. Giancotti Boscherini, *Ricerche lessicali su opere di Descartes e Spinoza*, 1969.
4. Eugenio Lecaldano, *Le analisi del linguaggio morale*, 1970.
5. Pierre Michaud-Quantin, *Etudes sur le vocabulaire philosophique du Moyen Age*, 1971.
6. Antonia Cancrini, *Syneidesis. Il tema semantico della "conscientia" nella Grecia antica*, 1970.
7. Marcella D'Abbiere, "Alienazione". In *Hegel. Usi e significati di Enta"usserung, Entfremdung, Vera"usserung*, 1970.
8. Alfonso Maierà, *Terminologia logica della tarda scolastica*, 1972.

9. René Roques, *Libres sentiers vers l'érigénisme*, 1975.
10. J.-R. Armogathe - J.-L. Marion, *Index des Regulae ad directionem ingenii de René Descartes*, 1977 (Corpus Cartesianum, vol. II).
13. L. Delatte - S. Govaerts - J. Denooz, *Index du Corpus Hermeticum*, 1977.
14. H. Usener, *Glossarium Epicureum, edendum curaverunt M. Gigante et W. Schmid*, 1977.
15. Marsilio Ficino, *Lessico greco-latino. Laur. Ashb. 1439*, a cura di Rosario Pintaudi, 1977 (Lexica Humanistica, vol. I).
- 16-17. Michele Ciliberto, *Lessico di Giordano Bruno*, 1979.
18. Giambattista Vico, *Principj di una scienza nuova. 1725*, vol. I, 1979.
19. Paolo Galluzzi, *Momento. Studi galileiani*, 1979.
- 20-21. *Ordo. Atti del II Colloquio Internazionale del Lessico Intellettuale Europeo*, a cura di M. Fattori e M. Bianchi, 1979.
22. Paolo F. Mugnai, *Segno e linguaggio in George Berkeley*, 1979.
- 23-24. Marta Fattori, *Lessico del Novum Organum di Francesco Bacone*, 1980.
25. Giambattista Vico, *Principj di una scienza nuova. 1725*, vol. II, *Concordanze e indici di frequenza a cura di A. Duro*, 1981.
26. *Res. III Colloquio Internazionale del Lessico Intellettuale Europeo*, atti a cura di M. Fattori e M. Bianchi, 1982.
27. U. Berni Canani - E. Brunet - R. Busa - G. Th. Guilbaud - J. Hamesse - Ch. Muller, *L'analisi delle frequenze. Problemi di lessicologia*, a cura di M. Fattori e M. Bianchi, 1982.
28. Andrea Bozzi, *Note di lessicografia ippocratica. Il trattato sulle arie, le acque, i luoghi*, 1982.

29. *Global Linguistic Statistical Methods to Locate Style Identities. Proceedings of an International Seminar. Gallarate June 5-7, 1981*, edited by R. Busa, 1982.
30. Anna Maria Bartoletti Colombo, *Lessico delle Novellae di Giustiniano*, vol. I (A-D), 1983.
31. Robert Halleux, *Indices Chemicorum Graecorum. I. Papyrus Leidensis. Papyrus Holmiensis*, 1983.

Rapporto di Franco RIZZI sulla missione di contatto e di informazione compiuta a Marseille presso il Centre de Mathématiques Sociales CNRS-EHESS e il Centre d'Analyse du Changement Economique et Social CNRS-EHESS, dicembre 1984

Sede: c/o Centre Pluridisciplinaire de la Vieille Charité, 2 rue de la Charité, 13002 Marseille.

Organizzazione: Il gruppo di ricerca è formato da: Jean Daniel Gronoff (programmatore analista), Alain Guillemin (sociologo), Pierre L. Jordan (antropologo).

Compiti: Questa équipe sta studiando le traiettorie familiari e l'evoluzione dei modi di trasmissione in una comunità rurale dell'Alta Provenza (Blègiers) tra il 1800 e il 1950. La ricerca si pone come scopo di capire come i gruppi familiari si riproducano ed evolvano e quali strategie mettono in atto a tal fine. Inoltre gli studiosi in questione stanno cercando di elaborare dei metodi di trattamento automatico dei dati che permettono di studiare congiuntamente le relazioni genealogiche, le strategie matrimoniali e le pratiche di trasmissione patrimoniale.

Macchine: I dati relativi a nascite, matrimoni e morti sono stati caricati su un Apple II. L'équipe lavora su macchine HEWLETT-PACKARD 498-45B.

Programmi: prodotti in loco con linguaggio Pascal.

Collaborazioni: Si prevede di organizzare un incontro a Roma con il gruppo di ricerca "Trattamento automatico delle genealogie" (cf. scheda a parte), al fine di verificare la possibilità di generalizzare un sistema di trattamento automatico delle genealogie.

Rapporto di Tito Orlandi sulla missione di contatto e di informazione compiuta a Milano (Facoltà di Lettere, CCA, CILEA). 19 aprile 1984

Le attività di ricerca umanistica attraverso il computer presso l'Università di Milano (Statale) si vanno organizzando intorno al "Polo di calcolo" della Facoltà di Lettere. Esso è a sua volta connesso con due organismi unificati di Università e di Regione: il Centro di calcolo automatico (dell'università) ed il CILEA (regionale). Attività analoghe sono svolte da tempo dalla Facoltà di Scienze politiche, con un suo Polo di calcolo assai attivo. Sembra però che questo tipo di attività, basato soprattutto su statistiche sociologiche, non possa rientrare, almeno per il momento, nell'ambito delle nostre ricerche.

1. Polo di calcolo della Facoltà di Lettere e Filosofia dell'Università degli Studi di Milano.

Sede: Presso il Dipartimento di Studi filosofici, situato nell'edificio principale della Facoltà (via Festa del Perdono 7, 20122 Milano. Tel. 02.8846.381). È stato messo a disposizione un locale.

Direzione e organizzazione: Non esiste ancora una direzione ufficiale, in quanto il Polo di calcolo è tuttora in via di organizzazione. Attualmente il responsabile è il prof. Mangione, ordinario di Logica e direttore del Dipartimento. Fra i collaboratori è opportuno ricordare il Dott. Edoardo Ballo. Finanziariamente il Polo

dipende dalla Facoltà nel suo complesso, ma è ancora incerto il sistema amministrativo concreto, in quanto la Presidenza non amministra attualmente fondi propri. Dal punto di vista ufficiale, esiste una decisione presa dal Consiglio di facoltà di costituire il Polo di calcolo; è stato in seguito anche proposto un regolamento, che tuttavia non è stato fino ad oggi discusso.

Note: Il regolamento proposto è stato fatto sul modello della "Bozza" in funzione nel nostro Servizio per l'informatica di Roma. Storicamente, il Polo è nato, per quanto mi è stato dato di capire, per impulso del Centro di calcolo automatico (vedi sotto), che anche ha provveduto a dotare gratuitamente il Polo di alcune macchine (PC IBM; terminale UNIVAC UTS20). Prima di costituire il Polo è stato fatto un censimento degli eventuali interessati, seguito da una riunione. Sono stati poi organizzati (sempre ad opera del CCA) dei corsi di "iniziazione". Un primo corso generico; poi corsi finalizzati ad alcune utilizzazione del Personal Computer: BASIC, elaborazione dati (DBII), composizione testi (WordStar).

Compiti: Offrire ad utenti della Facoltà di Lettere l'attrezzatura per automatizzare alcune ricerche.

Macchine: 1 Terminale UNIVAC UTS20, con relativo modem e multiplexer; inoltre una stampante. 1 Personal Computer (PC IBM). 1 Word Processor Olivetti, con relativa stampante.

Note: Si sta studiando un collegamento con Pisa, Istituto di Linguistica computazionale, forse attraverso una rete di personal computer (non ho capito bene il motivo, ma ho preferito non approfondire). Il terminale è collegato col mainframe del CILEA. Il CCA ha a Milano funzioni differenti (cf. sotto). Non vi sono progetti per usare personal computer come terminale intelligente. Non è stata

presa in considerazione l'idea di usare del Polo come intermediario-macchine fra utenti singoli e mainframe. Si sono avute difficoltà (peraltro superate nel giro di qualche giorno), a causa della rete telefonica, nel collegare il terminale col main-frame del CILEA.

Programmi: Quelli a disposizione presso il CILEA (cfr. sotto).

Utenti: La cattedra di Logica si serviva già del calcolatore per esperimenti di linguaggi logici ecc. Si sta organizzando l'automazione della biblioteca della Facoltà e di quelle degli istituti e dipartimenti tramite il programma BOMAS (cfr. sotto, CILEA). Nessun altro utente al momento ha progetti concreti.

Nota: A Milano si tende a non differenziare l'utenza "biblioteche" da quella "ricerche", come sta avvenendo a Roma. Ciò accade, ritengo, a causa del minor volume di lavoro prevedibile in ambedue i campi, e comunque mi sembra un errore. Tuttavia è un problema che potrebbe essere discusso utilmente.

Possibile collaborazione: È importante organizzare uno scambio continuo di informazioni; ma non si vedono per ora iniziative di ricerca concreta che offrano possibilità di collaborazione.

Nota: differente è la situazione del CILEA come tale: cfr. sotto.

Pubblicazioni: Nessuna.

2. Centro di calcolo automatico dell'Università degli Studi di Milano.

Sede: via Colombo 71, 20133 Milano (sede propria; edificio). Tel. 2361060 2361176.

Direzione e organizzazione: Direttore prof. Antonio Liverani - Consiglio Direttivo - Comitato Tecnico Scientifico. Il CCA è formalmente un Centro di servizi interdipartimentale. Ho parlato

con il prof. Paolo Monti, che si occupa in particolare delle relazioni con gli utenti "umanistici".

Nota: Il CCA è relativamente piccolo (a paragone, per esempio, col CCI di Roma) perché, come si è detto sopra, l'organizzazione generale relativa all'informatica si basa soprattutto sul CILEA.

Compiti: Consulenza ai dipartimenti ed istituti che hanno problemi di calcolo. I ragguagli dati sono di natura sia tecnica sia commerciale (prezzi ecc.). Le soluzioni proponibili possono essere: (a) per mezzo delle attrezzature del CCA; (b) su consiglio del CCA, con attrezzature in proprio; (c) per mezzo delle attrezzature del CILEA, eventualmente tramite interfaccia CCA.

Organizzazione di corsi di "alfabetizzazione" e di introduzione in vari campi dell'informatica, anche ad uso dei diversi poli di calcolo e facoltà. Favorire la nascita di poli di calcolo decentrati, anche con l'acquisto e cessione delle macchine necessarie ad avviarli.

Finanziare la didattica degli istituti che fanno tali tipi di applicazioni.

Iniziativa di valutazione di personal computer e di certificazione delle macchine. Prende contatto con le case, chiedendo che gli vengano inviate le macchine, che sono sottoposte a "prove severe".

Note: Il CCA favorisce, come politica generale, l'uso di Personal Computer come apparecchiature principali per i Poli di Calcolo. È in progetto l'acquisto di software per PC con possibilità giuridica di distribuzione ai Poli decentrati.

Macchine: Il mainframe è costituito da 2 PDP11/34. Tra poco però essi saranno sostituiti da un VAX. Il main-frame ha un programma di connessione col CILEA in modo che ogni terminale

collegato col PDP (poi col VAX) può collegarsi automaticamente col CILEA. Consuete periferiche, anche speciali (plotter etc.). Grappolo molto consistente di terminali incroni collegati con il mainframe del CILEA (non attraverso il mainframe del CCA). Sono sistemate in apposite sale, a disposizione degli studenti ecc. È in progetto una rete di collegamento fra tutti i calcolatori dei dipartimenti ed istituti per convogliarli verso il CILEA, con risparmio nelle tariffe telefoniche. È in progetto l'acquisto di un personal computer messo a punto dalla Hyphen, specialmente adatto per la fotocomposizione. È multifont, cioè compone caratteri speciali sia su schermo che in uscita (verso la macchina di fotocomposizione).

Programmi: nessun programma speciale d'interesse umanistico; per questo si suppone che gli utenti si appoggino al CILEA. Possiamo accennare al fatto che per le necessità di Scienze politiche è stato fornito il package di tipo statistico SPSS.

Utenti: La principale utenza del CCA, dal punto di vista delle sue proprie macchine, è oggi costituita dagli studenti di facoltà scientifiche, che preparano una tesi sul o col calcolatore. Per il resto, come si è detto, il CCA funziona piuttosto da interfaccia col CILEA o come consulente "teorico".

Possibili collaborazioni: scambio di informazioni. Il CCA stampa una "circolare informativa" trimestrale (cfr. sotto).

Pubblicazioni: Circolare informativa del Centro di Calcolo Automatico. Il Gruppo di ricerca IDU è stato messo nell'indirizzario. Da parte nostra si potrà organizzare un cambio. Il CCA non ha pubblicazioni che ne descrivono il funzionamento e l'organizzazione.

Nota: È opportuno, per questo e per casi consimili, iniziare la pubblicazione informale di un Bollettino o simili del nostro Gruppo

di ricerca, contenente, per esempio, estratti dai rapporti informativi e dai verbali delle nostre riunioni. Più tardi vi saranno, come previsto, i risultati della Banca Dati in via di formazione.

3. CILEA (Consorzio Interuniversitario Lombardo per la Elaborazione Automatica).

Sede: Via Raffaello Sanzio 4, 20090 Segrate/Milano. Tel. 02.2132541. Telex 31330 CILEAMI.

Direzione e organizzazione: Cons. di Amministrazione. Direzione. Ho parlato con l'ing. Giovanni Meloni.

Nota: Il colloquio è stato molto utile per l'estrema disponibilità a collaborare riscontrata. In particolare sono stati forniti parecchi manuali ecc. d'interesse generale, ai quali si farà riferimento in seguito. Essi sono a disposizione (al momento su richiesta, perché occorrerà organizzare in questo senso la biblioteca futura dello SPI).

Compiti: fornire tutti i servizi di informatica alle Università della Lombardia (Milano Statale, Bocconi, Politecnico, Cattolica; Pavia; Brescia. Non ancora consorziato ma associato Istituto universitario di Bergamo).

Nota: Un analogo consorzio è il CINECA, che serve 14 università del Nord-Est ecc. Il tentativo di formare un CINOCA per il Nord-Ovest è abortito.

Macchine: Il main-frame è un sistema UNIVAC 1100/82 (in affitto). Verrà presto sostituito da un 1100/90 (in leasing).

Fotocompositrice.

Unità UTS40 completa di sistema di dischetti ecc. (cfr. il manuale).

Altre usuali periferiche.

Note: La fotocomposizione è utilizzata mediante il package TEX-METAFONT. È in progetto l'acquisto di una stampante laser (meno costosa) per il lavoro di routine (bozze ecc.) prima del passaggio in fotocomposizione.

Collegamenti: È in funzione il collegamento EURONET. È in progetto il collegamento nazionale, che interesserà prima di tutto Roma. Sembra possa essere pronto in pochi mesi.

Programmi: Il software disponibile è elencato nel manuale. Ricordiamo tuttavia:

TEX per fotocomposizione, in particolare anche da file creato da Text-Editor, naturalmente con inseriti i necessari caratteri di controllo (cfr. il manuale).

BOMAS per la gestione delle biblioteche. È usato però anche da un gruppo di fisica presso la Biblioteca nazionale di Brera per la gestione dei rapporti scientifici e della corrispondenza ivi custoditi.

Nota: Il programma di Data Base di uso generale (cfr. manuale sul software disponibile) non sembra soddisfacente. Manca un programma di Concordanze. Il CILEA è concepito abbastanza strettamente come centro di servizi "di macchina". Il software di carattere particolare è lasciato piuttosto all'iniziativa degli utenti.

Utenti: Quelli di tipo umanistico passano attraverso il CCA (cfr. sopra).

Possibili collaborazioni: Per quanto mi è possibile giudicare, potrebbero essere utili specialmente nel campo della fotocomposizione e della lettura di dischetti IBM compatibili tramite UTS40, per la quale il CILEA ha già disponibili software ed esperienza. Inoltre ogni futuro contatto con le università della Lombardia passerebbe naturalmente attraverso il CILEA.

Pubblicazioni: Manuali d'interesse vario, ivi compreso linguaggio di controllo del 1100 ecc.

Osservazioni conclusive: In questa missione penso di aver avuto la possibilità di valutare la situazione non della sola Facoltà di Lettere di Milano-Statale, ma di tutta la Lombardia. Nel complesso sembra che, nonostante l'ambiente "strutture" sia favorevole allo sviluppo di iniziative di ricerca umanistica tramite calcolatore, pochissimi siano i ricercatori che lavorano in questo campo, e nulli i risultati concreti (un punto interrogativo rimane forse l'Università Cattolica e la relazione col padre Busa). È da prevedere che nei prossimi anni la situazione cambierà. L'attenzione sembra centrata principalmente sull'attività di Pisa (Zampolli).

Rapporto di Giovanni Adamo e Tito Orlandi sulla missione di contatto e di informazione compiuta a Oxford (Oxford University Computing Service - Computing in the Arts) il 19 febbraio 1985

Sede: Oxford University Computing Service (OUCS) - 13 Banbury Road – Oxford OX2 6NN - Tel.: (0865) 56721.

Direzione e organizzazione: Il settore "Computing in the Arts" fa parte dello User Services Group dell'OUCS. Section Manager: Prof. Susan Hockey. Programmer: Ian Marriott. Progr./Local Advisor: Catherine Griffin. External Advisor: Alastair Pearce. KDEM Supervisor: Grazyna Cooper. KDEM Operator: Dee Holl.

Orario di ricevimento: 8.30 - 17.30.

Compiti: Assistenza agli studiosi di discipline umanistiche che vogliono far uso dell'elaboratore, sia in maniera indipendente, che collegandosi al mainframe.

Attività: Da segnalare particolarmente l'organizzazione e la gestione di un "Text Archive", nel quale sono memorizzati testi di vari ambiti culturali e linguistici. L'Archivio raccoglie nastri prodotti da vari Centri di tutto il mondo. La lista delle opere memorizzate e le istruzioni per ottenerne copia sono depositate presso l'Archivio IDU. È stato creato un servizio di lettura ottica di testi a stampa tramite la "Kurzweil Data Entry Machine" (KDEM). Informazioni sul servizio e sulle modalità di accesso si trovano presso l'Archivio IDU. Vengono anche svolti corsi d'insegnamento teorici e pratici su: analisi di testi; applicazioni alla ricerca storica; programmazione in SPITBOL; uso della fotocompositrice Lasercomp (cfr. sotto).

Macchine: L'OUCS usa come mainframe un elaboratore ICL 2988 (sistema operativo VME) affiancato da 3 Digital VAX 11/780. Per l'output di stampa è da segnalare una macchina Monotype Lasercomp Phototypesetter che produce stampe di alta qualità

particolarmente adatte alla lavorazione off-set. Il settore umanistico gestisce in particolare il KDEM (cfr. sopra) e personal computer BBC e IBM e relative local printers.

Collegamenti: Alcuni dei gruppi di ricerca dell'università che hanno ottenuto consulenze dall'OUCS si sono poi equipaggiati con propri terminali o personal computer collegati con il mainframe.

Programmi: "Oxford Concordance Program" (OCP): programma completo di spogli lessicali. Le specifiche e le modalità per l'acquisizione si trovano presso l'archivio IDU.

"Multilingual Scribe": programma per IBM.PC prodotto dalla Gamma Productions (California) che consente di creare anche su video set di caratteri differenti e di produrre file formattati contenenti anche più set nello stesso tempo. Per l'output di stampa non si hanno ancora esperimenti soddisfacenti.

Possibili collaborazioni: Scambio di testi memorizzati e scambio di software.

Pubblicazioni: Opuscoli relativi alle attività del Centro (cfr. sopra).

Centro di elaborazione automatica di dati e documenti storico artistici, Scuola Normale Superiore di Pisa

Sede: Scuola Normale Superiore, Piazza dei Cavalieri 7, 56100 Pisa, Tel. 050.597111 / 597268

Direzione e organizzazione: Il CEADDSA è nato nel 1980 dalla collaborazione tra informatici, linguisti e storici dell'arte per la progettazione e gestione di banche dati relative ai beni culturali. Il Centro è diretto dalla prof.ssa Paola Barocchi, ordinario di Storia della critica d'arte. Gruppo di lavoro: Gabriella Cantini Guidotti, Barbara Cinelli, Laura Corti, Miriam Fileti Mazza, Franco Martorana, Umberto Parrini, Rossella Tarchi, Massimo Tosi. Consulenti: Lina Bolzoni, Nicoletta Calzolari, Giovanni Nencioni, Salvatore Settis, Lucia Tongiorgi.

Finalità: Sperimentazione di possibili programmi relativi ai beni culturali con particolare riguardo alle testimonianze documentarie, al lessico tecnico, al lessico storico, alle bibliografie, ecc. Tale sperimentazione è disponibile per l'utenza scientifica e il Centro si propone di offrire una consulenza e collaborazione ad Enti diversi. (Redatto dai responsabili del Centro).

Rapporto di Riccardo Duranti sui contatti con SISCO (Sistemi Cognitivi) s.r.l. di Roma

Sede: La SISCO ha un suo ufficio in Via F. Fiorentini, 106 - 00159 Roma, ma ha intenzione di trasferirsi in un prossimo futuro in via Morgagni. Telefoni: 433 982 (linea ordinaria) e 439 2167 (linea trasmissione dati).

Direzione e organizzazione: La SISCO ha la struttura di una s.r.l. i cui soci sono Giovanni Lariccia, Mauro Laeng, Stefania Panni e Stefano Lariccia. In pratica il lavoro è organizzato secondo un modello di laboratorio di ricerca e la direzione dei vari progetti è affidata di volta in volta ai responsabili del programma di ricerca.

Attività: La SISCO si definisce "un laboratorio per la conoscenza... che ha lo scopo di affrontare in modo integrato i problemi della cultura, della professionalità e della formazione nel campo delle Nuove Tecnologie dell'Informazione nella società postindustriale." In quest'ambito ha svolto numerose ricerche.

Macchine: Presso la sede della SISCO ho notato tre Personal IBM XT, di cui uno con schermo grafico, un APPLE II ed un APPLE MacIntosh, uno Zenith ed un Commodore 64 portatile. La loro biblioteca di software mi è parsa molto fornita.

Possibili collaborazioni: Stefano Lariccia, con cui ho personalmente parlato, si è dimostrato molto interessato alle attività dell'IDU, specialmente per uno scambio di informazioni bibliografiche e di esperienze di formazione professionale di tipo informatico per del personale di provenienza umanistica. Ha ventilato la possibilità di affidare a membri dell'IDU il compito di recensire programmi stranieri di interesse applicativo in campo umanistico, in vista di una loro eventuale traduzione in italiano.

Scheda informativa relativa a: Trattamento automatico delle genealogie a cura di Franco Rizzi

Sede: c/o Dipartimento di Studi Storici dal Medio Evo all'Età Contemporanea. Università

degli Studi di Roma "La Sapienza", P.le Aldo Moro 5, 00185 Roma.

Direzione: Prof. Franco Rizzi. L'équipe è formata per gli aspetti più strettamente informatici dalla prof. Mirella Schaerf (direttrice del Centro interdipartimentale di calcolo scientifico), dal dr. Antonio Magi Spinetti (CICS) e dal prof. Sergio Camiz (Dip. Di Ingegneria).

Compiti: Le attività di ricerca interessano prevalentemente l'ambito della storia sociale e si propongono di analizzare il ruolo della famiglia nella formazione dei circuiti di potere in alcune comunità rurali del Lazio tra XVIII e XX secolo. Inoltre il trattamento automatico delle genealogie permetterà di analizzare in maniera adeguata il complesso sistema matrimoniale e il rapporto tra questo e la trasmissione patrimoniale.

Macchine: c/o Centro interdipartimentale di calcolo scientifico dell'Università di Roma "La Sapienza": UNIVAC 1100/82.

Programmi: È in allestimento un programma specifico di trattamento dei dati matrimoniali. Nella prima fase le schede relative ai matrimoni sono state caricate in modo sequenziale in una Banca Dati di tipo relazionale: TROVADOC, dell'Università di Roma. La prima operazione di pulizia dei dati è stata fatta utilizzando le possibilità della banca dati stessa ricorrendo per i casi più difficili e incerti alla verifica manuale sui libri matrimoniali o di battesimi.

Rapporto di Annamaria Valente e Franco Bianchi sulla missione di contatto e di informazione compiuta a Tübingen, dicembre 1984

Sede: c/o Zentrum für Datenverarbeitung der Universität Tübingen, Abteilung fuer Literarische und Dokumentarische Datenverarbeitung, Brunnenstrasse 27, 7400 Tübingen

(RFT), tel. (0)7071-29-2901 (Segreteria)

Direzione e organizzazione: Direttore: Prof. M. GRAEF, Vice-direttore: Dott. Wilhelm Ott. Segreteria: G. Sichler. Collaboratori: H. Fuchs, D. Kottke, K. Schaekle, E. Serra.

Compiti: Il Centro di calcolo di Tübingen (= ZDV) rappresenta una istituzione centrale al servizio soprattutto dell'università di Tübingen, della Fachhochschule di Reutlingen e della Paedagogische Hochschule di Reutlingen. Fornisce un servizio relativamente ai seguenti settori: ricerca, insegnamento, amministrazione. Si propone di: 1) promuovere l'impiego della elaborazione di dati mediante corsi e altre misure di formazione e di perfezionamento, 2) di realizzare progetti di ricerca, 3) di prestare assistenza nella realizzazione di progetti fornendo: a) consulenza scientifica nella preparazione di progetti riguardanti l'elaborazione elettronica di dati, b) consulenza per programmi, c) creazione di una biblioteca di programmi, tramite programmi di utilizzo di interesse generale.

Macchine: SPERRY 1100/80 single processor; IBM 4341; BASF 7/88 (IBM compatibile).

Programmi: Viene usato il Sistema di Programmi TUSTEP (vedi descrizione sotto).

Nota: Ogni anno presso il Centro di Calcolo: 1) viene presentata una relazione sull'attività svolta; 2) si tengono dei Colloqui sull'applicazione dell'elaborazione elettronica nelle scienze

umanistiche (mesi di: febbraio, giugno, novembre); 3) viene tenuto un corso estivo della durata di circa 10 giorni, nel mese di settembre.

Utenti: Studiosi e ricercatori che ne facciano richiesta.

Descrizione del programma "TUSTEP": TUSTEP consiste in un insieme di programmi per mezzo dei quali è possibile eseguire le seguenti funzioni fondamentali:

- registrazione/correzione (online e batch)
- copia e manipolazione
- confronto di redazioni diverse
- scomposizione dei testi in elementi classificabili ed estrazione di parti caratteristiche
- preparazione di una classificazione di testi/di elementi test.
- raccolta in registri/indici di elementi test. classificati
- pubblicazione dei testi con diversi mezzi (stampante lineare provvista di una grande quantità di segni, stampante a margherita, COM, fotocomposizione).

A questi programmi se ne aggiungono altri di tipo organizzativo, per es. trasferimento di dati su altri impianti, utilizzazione di procedure, trattamento di errori ecc. I singoli programmi vengono chiamati sulla base di linguaggi di comando del rispettivo sistema operativo (non all'interno di un linguaggio di programmazione).

Campi di applicazione: documentazione, bibliografia, indici, registri, concordanze, edizioni, critica testuale, statistiche linguistiche, analisi metriche e stilistiche, analisi di fonti storiche, fotocomposizione.

Utenti: Studiosi e ricercatori che ne facciano richiesta [finora (dicembre 1984) richieste solo da parte di istituzioni della RFT].

Si noterà come fossero ancora rudimentali gli strumenti, soprattutto software, di cui potevano disporre gli umanisti. L'input era effettuato mediante terminali assai poco amichevoli. Gli editor testuali erano prodotti all'interno dei centri di calcolo e dunque erano fra loro incompatibili. Ma la riflessione sugli aspetti metodologici era molto vivace, a mio modo di vedere assai più di quanto non lo sia adesso. Essa tuttavia si manteneva nell'ambito delle singole applicazioni, e solo col tempo si farà strada l'esigenza di un apprezzamento metodologico trasversale.

Continuando la descrizione del lavoro compiuto dall'IDU, ricorderemo il monitoraggio delle liste e-mail e dei siti web relativi all'informatica umanistica, i cui risultati fanno parte dell'archivio del CISADU, che ci auguriamo non vadano perduti. Inoltre l'IDU pubblicava una collana (presso l'editore Bulzoni), che vide apparire:

Giovanni Adamo, *Bibliografia di informatica umanistica*

Raul Mordenti, *Informatica e critica dei testi*

Francesco Tiradritti (a cura di), *Informatica ed egittologia all'inizio degli anni '90*

Domenico Fiormonte (a cura di), *Informatica umanistica. Dalla ricerca all'insegnamento*

Nicola Tangari, *La codifica delle note. Analisi e diffusione dei dati musicali*

Tito Orlandi, *Per l'informatica nella Facoltà di Lettere*

Giuseppe Gigliozzi (a cura di), *Studi di codifica e trattamento automatico di testi*

Paola Moscati (a cura di), *Trattamento di dati negli studi archeologici e storici*

Giovanni Adamo (a cura di), *Trattamento, edizione e stampa di testi con il calcolatore*

Venendo ora all'attività didattica, essa obbedì al principio di un progressivo affrancamento da istituzioni e persone al di fuori della Facoltà di Lettere, salvo che per le discipline di carattere strettamente tecnico; inoltre di una profonda interdisciplinarietà metodologica fra l'informatica e i vari campi delle applicazioni umanistiche.

Diamo come esempio la conformazione organizzativa del Corso di perfezionamento "Informatica per le scienze umanistiche" dell'anno accademico 1990-91:

Università degli Studi di Roma - La Sapienza, Facoltà di Lettere e Filosofia, e Lessico Intellettuale Europeo del CNR, in collaborazione con: Dipartimento di Informatica e Sistemistica, Centro Interdipartimentale di Calcolo Scientifico, Servizio per l'Informatica della Facoltà di Lettere. Tito Orlandi (Direttore), Giorgio Ausiello, Mirella Casini Schaerf, Paolo Ercoli, Giuseppe Gigliozzi.

Seminari: Giovanni Adamo, Giovanni Azzena, Carlo Cellucci, Ivan Di Stefano, Maria Teresa Leoni Zanobini, Corrado Maltese, Paola Moscati, Annamaria Paci. [Ricordo che in anni precedenti avevano tenuto seminari Luigi Cerofolini e Ugo Berni-Canani].

Il Corso intende fornire tutte le conoscenze relative a quella parte dell'informatica che si occupa del trattamento di dati nell'ambito delle discipline umanistiche. Costituiranno oggetto del Corso i principi generali dell'informatica e le metodologie di applicazione nel settore umanistico.

Linee del programma:

1. Fondamenti di informatica per le discipline umanistiche:
 - a. Sviluppo storico dell'informatica e rapporti con le discipline umanistiche.

- b. Fondamenti delle applicazioni nel settore umanistico.
- 2. Principi di informatica:
 - a. Algoritmi e progettazione di algoritmi come teoria della gestione dell'informazione.
 - b. Linguaggi di programmazione procedurali.
- 3. Strutture logiche e informative:
 - a. Macchina di Von Neumann e struttura funzionale di un elaboratore.
 - b. Algebra relazionale e applicazioni.
 - c. Basi di dati e rappresentazione della conoscenza.
- 4. Sistemi di elaborazione:
 - a. Architettura dei sistemi di elaborazione concentrati e distribuiti.
 - b. Sistemi operativi, linguaggi evoluti, programmi.
 - c. Gestione di archivi
- 5. Analisi informatica di testi:
 - a. La codifica come modo per rappresentare un testo.
 - b. Indici, concordanze e lessici.
 - c. Basi di dati testuali e fattuali.
 - d. Lemmatizzazione e Dizionari di macchina.
 - e. Applicazioni statistiche, problemi di ecdotica, analisi dello stile, attribuzioni d'autore.
- 6. Teorie e modelli computazionali delle strutture letterarie:
 - a. Strutture linguistiche e sistemi formali.
 - b. Modelli semantici e strutture narrative.
 - c. Rappresentazioni di conoscenza e sistemi esperti.
 - d. Metodologie e applicazioni di Intelligenza Artificiale.
- 7. Prospettive:

Tavola rotonda conclusiva sui risultati del Corso e sulle prospettive di espansione delle applicazioni in campo umanistico.

Da questo Corso di perfezionamento si sviluppò, a partire dal 1994, un vero e proprio insegnamento, inserito nell'Ordine degli studi della Facoltà di Lettere e filosofia.

Ricorderemo l'organizzazione per l'anno accademico 2000-2001, che fu l'ultimo, in quanto la riforma universitaria che allora andava in vigore non ne permise la continuazione, per motivi che sarebbe troppo lungo qui riportare. L'insegnamento era così annunciato: Università di Roma La Sapienza - Informatica applicata alle scienze umane.

Insegnanti:

- Prof. Tito Orlandi (titolare per affidamento).
- Dott. Giuseppe Gigliozzi (ricercatore).
- Dott.ssa Paola Moscati (contratto integrativo).
- Dott. Giovanni Adamo (contratto integrativo "Lessicografia e lessicologia italiana").

Collaborazione tecnica:

- Dott. Alberto Camplani.
- Dott. Lanfranco Fabriani.

Biblioteca: è accessibile la biblioteca specializzata del CISADU. Supporto tecnico-scientifico: presso il CISADU. Esercitazioni: è accessibile (per quest'anno in regime di sperimentazione) l'aula informatizzata (piano IV), con posti di lavoro che verranno messi a disposizione degli studenti con modalità ed orario indicati nel cartello posto presso l'aula stessa. In alcune ore (verrà comunicata la tabella degli orari) saranno presenti dei collaboratori per aiutare gli

studenti che vorranno esercitarsi nei vari pacchetti insegnati nel Corso.

Il CISADU (Centro Interdipartimentale di Servizi per l'Automazione nelle Discipline Umanistiche) è situato presso la Facoltà di Lettere (piano seminterrato), e gestisce un'aula informatizzata per lezioni, esercitazioni, e utilizzazione singole; una biblioteca di informatica umanistica; il server per la rete di ricerca di Facoltà; un servizio di consulenza. È aperto tutti i giorni dalle 9 alle 13, e il lunedì e giovedì dalle 15 alle 17.

Il programma del corso, con quelle che potremmo chiamare dispense digitali è tuttora in linea presso il CISADU³ e mi risulta ancora visitato, anche se nessuno ne parla. Concluderemo col ricordare le proposte che, con varia fortuna, sono state fatte alle competenti autorità per inquadrare l'informatica umanistica negli ordinamenti universitari⁴. Esse hanno riguardato: l'istituzione di un raggruppamento disciplinare d'Informatica applicata alle discipline umanistiche; l'istituzione di un corso di perfezionamento in Lettere - Metodologie delle applicazioni multimediali; l'istituzione di una scuola di specializzazione in Metodi dell'informatica per le Scienze Umanistiche (unica proposta andata in porto).

La conclusione che si può trarre da quest'ultima esperienza è che in mancanza di un raggruppamento disciplinare specifico ogni altra iniziativa nel campo dell'ordinamento universitario così com'è ora strutturato è destinata al fallimento.

³ <http://rmcisadu.let.uniroma1.it/~orlandi/iasulez.html>.

⁴ Cfr. <http://rmcisadu.let.uniroma1.it/~orlandi/parodi/indice.html>.

Che cos'è, oggi, l'informatica umanistica?

L'impatto della tecnologia

Dino Buzzetti, Università di Bologna

1. Rispondere al quesito che ci è stato proposto, non può non tenere conto degli effetti che lo sviluppo delle tecnologie informatiche ha necessariamente esercitato sulle pratiche di ricerca che, da qualche decennio, hanno tratto ispirazione da alcune precise motivazioni teoriche. A proposito di tali presupposti teorici, una definizione dell'informatica umanistica, che può essere utilmente richiamata, si trova formulata, nella pubblicazione conclusiva di un *Thematic Network* europeo¹ dedicato allo *Advanced Computing in the Humanities* (ACO*HUM) e precisamente in un capitolo che tratta espressamente dell'applicazione di «metodi formali» alle discipline

¹ L'Agenzia Esecutiva per l'Istruzione, gli Audiovisivi e la Cultura (EACEA), istituita dalla Commissione delle Comunità Europee, così definisce gli obiettivi di un *Thematic Network*: «Lo scopo principale dei *Thematic Networks* è quello di elevare la qualità [...] in una determinata disciplina accademica o campo di studi», EACEA, *Erasmus Thematic Networks (Socrates): Objectives*, http://eacea.ec.europa.eu/static/en/overview/ThemNetwks_overview.htm (consultato il 23 marzo 2012).

umanistiche.

Nel testo, redatto sotto la direzione di Tito Orlandi, si legge:

[...] cercheremo di definire le proprietà essenziali di ogni tipo di informatica applicata sulla base del tradizionale binomio *strutture di dati* e *algoritmi*, riferendolo alle caratteristiche particolari di ciascuna disciplina: a) i metodi che servono a rappresentare l'informazione, in uno specifico campo di conoscenze, in modo tale da renderne possibile l'elaborazione mediante sistemi computazionali, producono le *strutture di dati* richieste da una certa disciplina; b) i metodi che servono per formulare i problemi e i procedimenti di ricerca, in uno specifico campo di conoscenze, in modo tale da permettere l'applicazione di procedure di elaborazione computazionali, producono gli *algoritmi* applicabili a una data disciplina.²

Se dunque l'informatica consiste nel rappresentare l'informazione con determinate strutture di dati ed elaborare l'informazione mediante algoritmi, quella particolare specie di *informatica applicata* in cui consiste l'*informatica umanistica* può essere considerata come informatica applicata alle discipline umanistiche e consisterà nei diversi modi di rappresentare l'informazione in forma idonea alle caratteristiche proprie delle varie discipline e nell'elaborare l'informazione con procedure idonee alle domande specifiche che la ricerca si pone nei diversi campi disciplinari.

² K. de Smedt et al. (ed.), *Computing in Humanities Education: A European Perspective*, Bergen, University of Bergen – HIT Centre, 1999, ch. 2, European studies on formal methods in the humanities, § 2.3, [M. Thaller,] *Defining humanities computing methodology*, <http://gandalf.aksis.uib.no/AcoHum/fm/fm-chapter-final.html> (consultato il 23 marzo 2012).

2. Assumendo questi presupposti teorici come una caratterizzazione effettiva dell'informatica umanistica, così come si è venuta sviluppando fin dal primo apparire dei suoi concreti progetti di ricerca, ci si può allora chiedere se la pratica comune delle applicazioni dell'informatica alle discipline umanistiche, alla quale abbiamo assistito, si possa riconoscere in questa sua definizione, nelle diverse fasi di sviluppo delle tecnologie informatiche, che di volta in volta si sono rese disponibili. In questa prospettiva si possono distinguere tre momenti principali dello sviluppo tecnologico, tali da influenzare in forma diversa l'adozione di metodi computazionali nelle discipline umanistiche.

Già al tempo dei grandi calcolatori, o *mainframe*, si è assistito al varo di grandi progetti di ricerca, soprattutto nel campo della linguistica computazionale, e alla creazione di importanti archivi linguistici e testuali. L'onerosità e la scarsa accessibilità dei dispositivi e delle attrezzature necessarie alla ricerca rendeva possibile solamente l'avvio di grandi progetti sostenuti istituzionalmente da rilevanti risorse finanziarie e condotti a termine con la partecipazione da ampi gruppi di studiosi qualificati. Si sono così costituiti i grandi centri di ricerca e le grandi banche dati per l'elaborazione di risorse soprattutto testuali.

Una fase nuova si è aperta con l'introduzione dei piccoli calcolatori, i cosiddetti *mini* calcolatori o *personal computer*.

L'accessibilità agli strumenti informatici è stata così offerta potenzialmente a tutti i ricercatori e si è assistito ad una grande proliferazione di iniziative individuali, o promosse da piccoli gruppi di ricerca. L'introduzione del *personal computer* ha anche comportato l'adozione di forme standard di codificazione primaria dell'informazione e l'interscambiabilità dei dati e dei programmi. Capacità di elaborazione potenziate potevano essere realizzate con l'impiego di

calcolatori di dimensioni intermedie, o *workstation*, che restavano però escluse dalle prime reti telematiche che collegavano, con protocolli ancora proprietari, solo i grandi calcolatori.

Un ulteriore incremento dell'accessibilità dei dati e della diffusione delle iniziative di ricerca si è prodotto con l'introduzione del *WorldWideWeb*. Con la sua rapida e immediata espansione, la pervasività del mezzo informatico per la comunicazione dei dati e delle informazioni ha praticamente coinvolto, a livelli diversi, ogni forma di attività e di produzione culturale. L'accessibilità e la diffusione di contenuti multimediali, alla cui prima apparizione si era già assistito nella fase precedente, ha reso possibile l'estensione dell'applicazione di procedure computazionali a materiali non esclusivamente testuali anche nel campo della ricerca umanistica.

3. Il succedersi di queste diverse fasi nella disponibilità degli strumenti offerti dalla tecnologia non è stato però privo di conseguenze sulla pratica effettiva dell'applicazione di metodi informatici in campo umanistico. Possiamo cercare di illustrarne la portata con un esempio tratto dal campo della filologia e della critica testuale. A chi avesse cominciato ad occuparsi di edizioni digitali fin dalla fine degli anni ottanta, si offriva allora la possibilità di associare le trascrizioni dei testi alle immagini delle fonti manoscritte solo attraverso la disponibilità di potenti *workstation* in grado di offrire risorse di calcolo sufficienti all'elaborazione di immagini digitali ad alta risoluzione e tali da permettere una riproduzione adeguata dei documenti originali. Ma nei sistemi a quel tempo disponibili³, la possibilità di associare singole parti della trascrizione alle porzioni

³ Cfr. M. Thaller, "Kleio: Ein Datenbanksystem", St. Katharinen, Max-Planck-Institut für Geschichte. In *Kommission bei Scripta Mercaturae*, Verlag, 1989.

corrispondenti delle immagini dei manoscritti da cui erano ricavate, presupponeva l'organizzazione del testo in elementi di informazione strutturata e l'integrazione di dati visuali in unico sistema di gestione dell'informazione, o *DataBase Management System* (DBMS).

L'edizione era concepita come un database⁴ e, ciò che più importa ai fini della nostra trattazione, la trascrizione era concepita non più come un mezzo per riprodurre il documento originale, alla stregua di una tradizionale trascrizione diplomatica, utile soprattutto al fine di trasmettere al lettore «un'idea più fedele della natura della fonte» manoscritta,⁵ ma come un dato fornito al computer «per essere ulteriormente elaborato»; così, in questa fase, la trascrizione

diventa un'attività di modellizzazione e di codifica dei dati, per estrarre dal manoscritto quanta più informazione possibile e per ricavarne analiticamente nuovi risultati. Da questo punto di vista, sia l'immagine che la trascrizione non sono più considerate come riproduzioni *fisiche* utili a ripresentare il documento originale, ma piuttosto come dati analitici miranti ad una nuova rappresentazione *logica* della fonte.

Mi sono deliberatamente riferito ad uno scritto di quegli anni, per mostrare come già nella fase d'uso del personal computer e fin nei primi tempi dell'espansione del web, la preoccupazione

⁴ D. Buzzetti, P. Pari, A. Tabarroni, "Libri e maestri a Bologna nel XIV secolo: Un'edizione come database". In *Schede umanistiche*. n.s., 6, 1992(2), p. 163-169 e D. Buzzetti, "Masters and Books in 14th-century Bologna: An edition as a database". In F. Bocchi, P. Denley (eds.), *Storia & Multimedia, Proceedings of the Seventh International Congress of the Association for History and Computing*. Bologna, 29 August-2 September 1992, Bologna, Grafis Edizioni, 1994, p. 642-646.

⁵ D. Buzzetti, "Image Processing and the Study of Manuscript Textual Traditions". In *Historical Methods*. 28, 1995(3), p. 145.

metodologica fosse ancora legata al lascito dell'esperienza maturata nel periodo precedente, contraddistinto dall'uso dei mainframe. L'insistenza sulla produzione di dati strutturati e la contrapposizione tra l'uso *logico* e l'uso *analogico* della rappresentazione digitale mettono in luce la permanenza di un'attenzione rivolta prevalentemente all'elaborazione dei dati e ad una concezione della rappresentazione dell'informazione funzionale soprattutto all'applicazione di procedure computazionali miranti all'elaborazione del contenuto informativo della fonte.

4. Mette conto insistere su questo aspetto perché fa risaltare, per converso, una tendenza prodottasi paradossalmente nel passaggio dai mainframe ai personal computer, per giungere infine all'attuale predominanza dei servizi web. La paradossalità del fenomeno consiste nello spostamento dell'enfasi dalla funzione che abbiamo definito "logica" alla funzione che abbiamo contraddistinto, metaforicamente, come "analogica" della rappresentazione digitale, che non è più vista, prevalentemente, come funzionale all'elaborazione del contenuto informativo del documento o dell'oggetto fisico rappresentato, ma che viene sempre più considerata come un mezzo per la riproduzione virtuale delle loro caratteristiche fisiche e materiali.

Nel campo dell'informatica umanistica questa tendenza al distacco dall'elaborazione del contenuto si è manifestata fin dalla fase di passaggio all'uso dei personal computer. Con l'affermarsi dei word processor e dei sistemi WYSIWYG (*What You See Is What You Get*) per la produzione dei documenti, nel campo stesso dell'infor-

matica, alla cosiddetta *data processing* o *database community* si è affiancata con sempre maggior peso la cosiddetta *document community*⁶.

Mentre l'interesse prevalente della *document community* mirava sostanzialmente a preservare l'invarianza del documento – e ciò valeva in particolare per i documenti di studio, oggetto della ricerca umanistica – l'interesse della *database community* mirava invece a preservare l'invarianza della semantica di dati. Così, mentre la *database community* «sceglieva di standardizzare la semantica» e quindi si preoccupava principalmente dell'elaborazione del *contenuto* dei dati, senza prestare particolare attenzione alla loro forma di *rappresentazione*, la *document community* «sceglieva di standardizzare la rappresentazione dei dati» e di generalizzare il *markup*, introducendo lo standard SGML e favorendo la marcatura non solo degli «aspetti tipografici dei documenti», ma anche di «proprietà più generali dei testi».

Le conseguenze di questi diversi orientamenti furono decisive nelle scelte delle comunità di ricerca che si dedicavano all'elaborazione automatica dei testi. Le tre associazioni scientifiche che li rappresentavano, la *Association for Computational Linguistics* (ACL), la *Association for Literary and Linguistic Computing* (ALLC) e la *Association for Computers and the Humanities* (ACH) decisero di promuovere la *Text Encoding Initiative* (TEI) e di adottare lo standard SGML per la marcatura dei testi. Tuttavia, benché il problema della semantica dei dati «non fosse irrilevante per la *document community*», sembrava che la definizione di una semantica per il contenuto dei

⁶ D. Raymond et al., "From Data Representation to Data Model: Meta-semantic issues in the evolution of SGML". In *Computer Standards & Interfaces*. 18, 1996, p. 25-36.

documenti «fosse un problema di difficile soluzione». Allo stesso modo, anche «i tentativi di definire una semantica nella comunità degli studiosi, e particolarmente nella Text Encoding Initiative, incontrarono molte resistenze». Così «la via indicata dallo SGML» sembrò allora la più «ragionevole» e si scelse

[di] insistere sul concetto dell'indipendenza dalle macchine e dai programmi e [di] fornire la base su cui in seguito si sarebbe potuta sviluppare la semantica, evitando però di proporre effettivamente una particolare specificazione⁷.

La conseguenza fu quella di spostare l'attenzione dall'elaborazione del contenuto alla mera rappresentazione dei dati. Delle due fondamentali questioni di cui si occupa l'informatica, e con essa l'informatica umanistica, la rappresentazione e l'elaborazione dell'informazione, l'elaborazione del contenuto è passata sempre più in secondo piano. L'improvvisa espansione del web e la possibilità di presentare sul web contenuti multimediali hanno ulteriormente accentuato questa tendenza.

5. I linguaggi su cui si fonda l'architettura del web, lo HTML e ora sempre più lo XML, sono, non a caso, linguaggi per la rappresentazione dei dati. Sono linguaggi che esprimono la struttura della rappresentazione, non la struttura di ciò che tale rappresentazione rappresenta; in altri termini, tali linguaggi non sono in grado di esprimere la struttura del contenuto informativo della rappresentazione, se non quando le due strutture siano isomorfe e possano essere poste in corrispondenza biunivoca. Di conseguenza, anche le

⁷ Ivi, p. 28.

elaborazioni applicabili alle risorse web, sono in gran parte dipendenti dai formati dei documenti, ossia dalla struttura della rappresentazione dei contenuti informativi che essi veicolano. Le possibilità di elaborazione dei documenti digitali accessibili sul web dipendono dalla struttura che tali linguaggi assegnano ai documenti, ossia dai vincoli imposti da una struttura gerarchica ad albero. In effetti, il linguaggio usato per elaborare dati in formato XML, lo XSLT, «assume in ingresso (*input*) una struttura ad albero e genera in uscita (*output*) un'altra struttura ad albero»⁸. Esso è uno dei componenti della famiglia di linguaggi XSL (*Extensible Stylesheet Language*), introdotti «per definire la trasformazione e la presentazione di documenti XML»⁹, e se può essere certamente considerato «un linguaggio di programmazione completo», viene principalmente impiegato «per trasformare i dati d'origine nella forma di presentazione voluta». In realtà lo XSLT può servire anche per produrre «dati di ingresso (*input*) per un'altra applicazione»¹⁰, ma in ogni caso trasforma i dati d'origine XML in altre strutture di dati organizzate gerarchicamente ad albero; esso conserva così la struttura tipica del formato dei documenti digitali che veicolano l'informazione e non può dare risposta all'esigenza di elaborare la struttura propria del loro contenuto informativo.

⁸ M. Kay, *What Kind of Language is XSLT? An analysis and overview*, <http://www.ibm.com/developerworks/library/x-xslt> (consultato il 23 marzo 2012).

⁹ W3C, *The Extensible Stylesheet Language Family (XSL)*, <http://www.w3.org/Style/XSL> (consultato il 23 marzo 2012).

¹⁰ M. Kay, *What Kind of Language is XSLT?*, cit.

Il web, d'altronde, nella sua forma originaria, era stato concepito per permettere l'accesso a distanza e la visualizzazione dei documenti ed è proprio per questa sua caratteristica che molti hanno potuto considerarlo come una potenziale biblioteca universale e che alcuni progetti miranti esplicitamente a questo scopo abbiano potuto essere intrapresi¹¹. Spesso, tuttavia, è stato osservato che tali progetti hanno mirato più alla quantità che all'accuratezza e si è rafforzata l'opinione che l'«ossessione per i numeri debba cedere il passo alle preoccupazioni per la qualità»¹². In effetti, ciò che viene prodotto da tali progetti di «digitalizzazione di massa» non è altro che «l'immagine di una copia del libro, una fotografia digitale delle sue pagine, associata a funzioni di ricerca molto limitate»¹³, che operano sui dati spesso imprecisi forniti da una lettura ottica automatica e priva di ulteriori controlli. Sicché, pur con l'aggiunta di un corredo essenziale di metadati, queste «enormi quantità di testo» vengono presentate «con un valore aggiunto minimale», se paragonato a quello delle edizioni digitali vere e proprie, in cui «piccole quantità

¹¹ Degni di nota «tre grandiosi tentativi», quali il Million Books Project, promosso nel 2002 dalla Carnegie Mellon University e chiamato anche Universal Digital Library (<http://www.ulib.org>, consultato il 27 marzo 2012); il Google Books Search, proposto nel 2002 da Larry Page, uno dei due fondatori di Google, e cominciato nel 2004 (<http://books.google.com>, consultato il 27 marzo 2012); e il MSN Book Search, della Microsoft, annunciato nel 2005 e cessato nel 2008: cfr. M. Deegan, K. Sutherland, *Transferred Illusions: Digital technology and the forms of print*, Ashgate, Farnham/Burlington, VT, 2009, p. 136 e sgg.

¹² J.-N. Jeanneney, *Google and the Myth of Universal Knowledge: A view from Europe*, Chicago, University of Chicago Press, 2007, p. 78.

¹³ M. Deegan, K. Sutherland, *Transferred Illusions* cit., p. 146.

di dati testuali codificati» sono pubblicati «con un considerevole valore aggiunto», fatto di testimoni diversi del testo, annotazioni, immagini, glossari e link esterni, organizzati in ambienti di lavoro (*work-sites*) complessi e dotati di sofisticate funzionalità¹⁴.

Di conseguenza, tali grandi progetti di digitalizzazione sono stati accolti con molte perplessità. Deegan and Sutherland hanno osservato che «ponendo mente alla ricerca», spesso non ci si guadagna altro che l'accesso a «contenuti superati, giustamente trascurati e inaffidabili» e si finisce col «risuscitare e riciclare un sapere stantio e non più attuale»¹⁵. Inoltre, esse fanno notare che «lo spazio informativo (*information space*) di Internet è di tipo completamente diverso da quello di una biblioteca»¹⁶, sicché ora «il paradigma della biblioteca universale è quello di Internet e niente affatto quello di una biblioteca»¹⁷. In questo spazio informativo aperto non vengono trasferiti i principi di organizzazione dell'informazione propri della tradizione biblioteconomica e quindi «si perde uno dei maggiori benefici che le biblioteche recano a quello spazio quasi senza confine che è costituito dalla nostra cultura erudita: l'ordine»¹⁸. Infine, sempre dal punto di vista della ricerca, il tipo di accesso ai testi imposto dal web rischierebbe di compromettere la capacità di interagire criticamente con essi, significativamente e in profondità, al punto da far sorgere una nuova forme di apprensione. Questa

¹⁴ Ivi, p. 142.

¹⁵ Ivi, p. 145.

¹⁶ Ivi, p. 150.

¹⁷ Ivi, p. 151.

¹⁸ Ivi, p. 149

preoccupazione affatto nuova (*new anxiety*) risiede in questo: è possibile che l'accesso istantaneo all'informazione, l'elaborazione automatica del testo e quel genere di stimoli visivi e di immedesimazione con lo schermo (*screen absorption*) che sostituiscono la lettura compro-mettano l'esercizio delle preziose capacità riflessive proprie del processo cerebrale della lettura (*reading brain*)?¹⁹

Si tratta semplicemente dell'inquietudine per il nuovo o di un problema reale, che modifica le nostre stesse attitudini mentali?

Si può argomentare che i grandi progetti di digitalizzazione non sono stati ideati come applicazioni proprie dell'informatica umanistica, ma che nascono più specificamente dalle soluzioni che la moderna tecnologia è in grado di offrire per facilitare l'accesso al materiale librario. Infatti, di *libri* si tratta e non propriamente di testi: il libro è solo il «veicolo fisico» del testo²⁰, l'«architettura dell'informazione»²¹ che esso contiene, e come Deegan e Sutherland non mancano di far notare, il progetto «Google Book Search (si noti – dicono – il corsivo aggiunto) non mette a nostra disposizione del testo elettronico, ma mette a nostra disposizione dei libri»²². Tuttavia, anche le edizioni digitali e altri veri e propri prodotti della ricerca nel campo dell'informatica umanistica sono stati condizionati

¹⁹ Ivi, p. 173. Il riferimento di Deegan e Sutherland al *reading brain* rimanda al saggio di M. Wolf, *Proust and the Squid: The story and science of the reading brain*, Cambridge, Icon Books, 2008, in cui si mette in risalto la funzione essenziale e la portata evolutiva della lettura nello sviluppo delle capacità mentali della specie umana.

²⁰ Ivi, p. 91.

²¹ B. Cope, A. Phillips (ed.), *The Future of the Book in the Digital Age*, Oxford, Chandos Publishing, 2006, p. 8.

²² M. Deegan, K. Sutherland, *Transferred Illusions* cit., p. 147.

dalle nuove funzionalità che il *Web* ha permesso di utilizzare, tanto che l'impatto delle nuove tecnologie di rete ha prodotto effetti quasi paradossali, che hanno agito in controtendenza rispetto allo sviluppo dei metodi computazionali applicabili alle discipline umanistiche. La comparsa del testo elettronico sul web sembra aver favorito in misura maggiore la «simulazione di specifici prodotti di stampa»²³, le cui «caratteristiche libresche», anche se non nella forma dell'«oggetto fisico», sono state «conservate in modo trionfale»²⁴, senza peraltro contribuire a stimolare ed estendere la nostra interazione critica col testo. A giudizio delle due studiose, il «pensare in profondità» è strettamente connesso col «leggere in profondità», ma «né l'una né l'altra attività viene agevolata dall'impazienza e dall'applicazione ininterrotta delle funzionalità offerte dalla rete – l'accesso istantaneo, i collegamenti multipli (*multi-linking*), l'elaborazione simultanea (*multi-tasking*) e la ridotta interazione con i diversi spazi informativi»²⁵. E come ci si può chiedere se «gli strumenti di *text mining* possano effettivamente servire a dar senso alla letteratura»²⁶, allo stesso modo, nel campo delle edizioni digitali, si può pensare che «le potenzialità del *computer* come strumento di visualizzazione abbiano prevalso sui suoi usi propriamente 'computazionali'»²⁷. Così, l'attenzione è stata rivolta soprattutto ai problemi della rappresentazione del testo, trascurando o ponendo in secondo piano lo sviluppo di procedure automatiche per l'elabo-

²³ Ivi, p. 27.

²⁴ Ivi, p. 116-117.

²⁵ Ivi, p. 180.

²⁶ Ivi, p. 176.

²⁷ Ivi, p. 75.

razione del suo contenuto informativo. Insomma, la rappresentazione dell'*espressione*, o della materialità del testo, ha prevalso sull'*elaborazione* del suo *contenuto*, ovvero dell'informazione che esso trasmette, e sebbene la rappresentazione del testo materiale sia costituita da un testo digitale, la *rappresentazione* del testo materiale ha prevalso sull'*elaborazione* dell'informazione contenuta nell'oggetto digitale che lo rappresenta.

6. È possibile cercare di comprendere questa situazione apparentemente paradossale e di rispondere in modo adeguato alla *new anxiety* avvertita dai critici più attenti della portata innovativa delle nuove tecnologie e del loro impatto sullo sviluppo delle nostre pratiche culturali? L'indizio di un nuovo punto di vista da cui considerare la questione può essere colto nell'uso, apparentemente non intenzionale, di particolari metafore riferite allo scambio di informazioni in rete. Esse illustrano in modo incisivo la fase di passaggio dallo scambio bilaterale e diretto dell'informazione tra sistemi isolati alla diffusione distribuita dell'informazione nel nuovo sistema informativo integrato costituito dal web.

Questo passaggio può essere descritto come un passaggio dal *Gopher* a *Lynx*. Il *Gopher* (1991) era un protocollo progettato per distribuire, ricercare e recuperare documenti in Internet facendo uso di server appositamente predisposti. *Lynx* (1992) era un *web browser* per terminali funzionanti in modalità carattere e privi di interfaccia grafica, in grado di supportare diversi protocolli, tra cui il *Gopher* e il protocollo HTTP che era stato introdotto per lo scambio di dati nel *WorldWideWeb*. È evidente che ci troviamo, con ciò, in una fase di passaggio tra la sperimentazione di sistemi isolati e il successo del nuovo sistema distribuito per la diffusione dell'informazione. L'uso metaforico dei termini, *gopher* e *lynx*, si fonda, rispettivamente,

sull'omofonia con le espressioni 'go for' (o 'go far') e 'links,' che alludono al modo di operare dei *gopher* e alla navigazione nel web attraverso i collegamenti (*links*) ipertestuali. È singolare però che si tratti di metafore biologiche che fanno leva sul nome di due specie animali, un roditore della famiglia dei geomidi (il *gopher*) e la lince (*lynx*), un genere della famiglia dei felini. Ma le metafore biologiche con le specie animali richiamano l'idea dell'evoluzione e proprio in questi termini viene portato a ragionare chi trova scritto che «il protocollo *Gopher* è ancora usato da alcuni entusiasti» e che «rimane ancora una piccola *popolazione* di *server* mantenuti in funzione». ²⁸

7. Che cosa ci permette di cogliere, a proposito del nostro problema, una prospettiva di tipo evolutivo? In primo luogo l'importanza dello sviluppo delle *funzioni* dell'organismo –nel nostro caso, dell'oggetto digitale – insieme all'evoluzione del suo *aspetto*. Se dunque, da una parte, gli oggetti digitali accessibili sul web possono essere visualizzati e presentati come riproduzioni fedeli di oggetti fisici corrispondenti, dall'altra, essi possono essere valutati sulla base delle possibilità di elaborazione del loro contenuto informativo. Da questo punto di vista, si possono riscontrare sviluppi convergenti in campi di ricerca diversi e interconnessi.

Nel campo dell'informatica umanistica si sta affermando la consapevolezza che i suoi metodi e le sue procedure convergono con quelli propri della rappresentazione della conoscenza, praticati nel campo dell'intelligenza artificiale:

²⁸ Cfr. "Gopher (protocol)". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, [http://en.wikipedia.org/wiki/Gopher_\(protocol\)](http://en.wikipedia.org/wiki/Gopher_(protocol)) (consultato il 5 aprile 2012).

Per l'informatica umanistica, la rappresentazione della conoscenza è un modo convincente, illuminante, e produttivo di fare ricerca umanistica – e per molti versi è ciò che l'informatica umanistica è venuta facendo, implicitamente, per anni.²⁹

John Unsworth, che ne è l'autore, trae questa affermazione da un confronto tra le funzioni svolte dalla rappresentazione della conoscenza nella pratica dell'intelligenza artificiale e le diverse forme di rappresentazione adottate nel campo dell'informatica umanistica. Alla domanda su che cosa sia l'informatica umanistica, Unsworth risponde nel modo seguente:

L'informatica umanistica è una pratica di rappresentazione, una forma di modellizzazione o, come afferma Wallace Stevens, di imitazione. È anche (come Davis e i suoi coautori affermano) un modo di ragionare e un insieme di assunti ontologici, e la sua pratica di rappresentazione è stata determinata dal bisogno di computazione efficiente, da un lato, e di comunicazione umana dall'altro.³⁰

Questa caratterizzazione dell'informatica umanistica è ricavata dall'analoga risposta alla domanda su che cos'è la rappresentazione della conoscenza in intelligenza artificiale che Randall Davis, Howard Shrobe e Peter Szolovits propongono «nei termini di cinque

²⁹ J. Unsworth, "Knowledge Representation in Humanities Computing", presentato a *eHumanities*, NEH Lecture Series on Technology & the Humanities, Lecture I, Washington, DC, April 3, 2001, <http://people.lis.illinois.edu/~unsworth/KR> (consultato il 5 aprile 2012).

³⁰ Id., "What is Humanities Computing and What is Not?". In *Jahrbuch für Computerphilologie*, 4, 2002, p. 71-84, <http://computerphilologie.tudarmstadt.de/jg02/unsworth.html> (consultato il 5 aprile 2012).

importanti e distintamente diversi ruoli che una rappresentazione svolge» e secondo la quale la rappresentazione può essere considerata come (a) «un surrogato», ovvero un modello – dice Unsworth – dell'oggetto rappresentato, (b) «una teoria frammentaria di ragionamento intelligente», (c) «un insieme di assunzioni ontologiche», (d) «un mezzo (*means*) per una computazione pragmaticamente efficiente» e, infine, (e) «un mezzo (*medium*) di espressione umana». ³¹

Come si può osservare, le forme di rappresentazione praticate dall'informatica umanistica sono qui definite in funzione dell'efficienza computazionale. Rappresentazione ed elaborazione dell'informazione non restano quindi separate, ma la forma di rappresentazione risulta progettata in funzione dell'applicazione di procedure computazionali volte a produrre risultati utili sul piano dell'elaborazione dell'informazione. Le conseguenze di questa concezione sono tratte dallo stesso Unsworth in un articolo presentato nel gennaio 2003 ad un convegno sul tema *Transforming Disciplines: Computer Science and the Humanities*, dove scrive:

Noi dobbiamo (dobbiamo ancora) dimostrare l'utilità di tutti i materiali che abbiamo digitalizzato nel corso degli ultimi dieci anni ed oltre – e dimostrarne l'utilità non solo per quanto riguarda la maggiore facilità di accesso, ma in modo più preciso, per quanto riguarda *ciò che ne possiamo fare* quando li abbiamo a nostra disposizione [...] ³²

³¹ R. Davis, H. Shrobe, P. Szolovits, "What is a Knowledge Representation?". In *AI Magazine*. 14, 1993(1), p. 17-33.

³² J. Unsworth, "Tool-Time, or 'Haven't We Been Here Already?' Ten Years in Humanities Computing", presentato al convegno *Transforming Disciplines: The*

Si tratta di un «invito evidente ad andare 'oltre il rappresentare'»³³, come recita esplicitamente il titolo di uno dei suoi articoli pubblicati successivamente, dove dichiara in modo ancora più diretto:

Abbiamo speso una generazione costruendo furiosamente biblioteche digitali e sono sicuro che ora costruiremo strumenti da usare in quelle biblioteche altrettanto furiosamente, per almeno un'altra generazione, e non vedo l'ora che ciò si faccia. Sono sicuro che il testo non verrà dimenticato, mentre saremo occupati nella costruzione dei nostri strumenti – ma sono anche certo che i nostri strumenti ci porranno in relazione con i testi in forme del tutto nuove.

Nello stesso saggio, Unsworth coglie distintamente il succedersi delle diverse fasi di sviluppo dell'informatica umanistica, che qui sono state messe in relazione con il parallelo sviluppo della tecnologia e con l'introduzione di importanti innovazioni tecnologiche. Così, infatti, egli afferma:

siamo, credo, all'inizio di quella che mi sembra la terza fase principale dell'informatica umanistica (*humanities computing*), che è passata dagli strumenti, negli anni '50, '60 e '70, alle fonti primarie, negli anni '80 e '90, e ora sembra ritornare di nuovo agli strumenti.

Humanities and Computer Science, Washington DC, 17-18 gennaio 2003, <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html> (consultato il 5 aprile 2012).

³³ D Buzzetti, "Oltre il rappresentare: Le potenzialità del markup". In *La macchina nel tempo: Studi di informatica umanistica in onore di Tito Orlandi*, a cura di L. Perilli e D. Fiormonte, Firenze, Le Lettere, 2011, p. 40.

Sicché, continua, «credo che stiamo arrivando ad un momento in cui la forma di attenzione che prestiamo alle fonti primarie si stia spostando dalla digitalizzazione all'analisi, [...] dalla rappresentazione all'astrazione», e non potremmo trovare una formulazione più chiara del passaggio dalla preoccupazione prevalente per la rappresentazione delle fonti a quella per l'elaborazione del contenuto.³⁴

8. Ed è proprio dal punto di vista dell'elaborazione del contenuto informativo degli oggetti digitali che si realizza la convergenza dell'informatica umanistica con le tendenze in atto in altri e diversi campi contigui di ricerca. Così nel campo specifico dell'organizzazione della conoscenza e della catalogazione a soggetto, Vanda Broughton, esperta di classificazione a faccette e di costruzione di thesauri, afferma:

Il lavoro di collaborazione attualmente in corso con studiosi operanti nel campo dell'informatica umanistica suggerisce che l'uso combinato di metodi di analisi a faccette e metodi di codifica del testo possono offrire una soluzione per migliorare la praticabilità degli strumenti di metadattazione e fornire mezzi più ingegnosi e sofisticati per la rappresentazione dei soggetti.³⁵

³⁴ J. Unsworth, "Forms of Attention: Digital Humanities Beyond Representation", presentato al terzo convegno del Canadian Symposium on Text Analysis (CaSTA), *The Face of Text: Computer-Assisted Text Analysis in the Humanities*, McMaster University, 19-21 novembre 2004, <http://www3.isrl.uiuc.edu/~unsworth/FOA> (consultato il 6 aprile 2012).

³⁵ V. Broughton, Finding Bliss on the Web: Some problems of representing faceted terminologies in digital environment, in C. Gnoli, F. Mazzocchi (ed.), "Paradigms and Conceptual Systems". In *Knowledge Organization, Proceedings of the Eleventh*

Ed è importante osservare che la convergenza si realizzi proprio sul piano dell'analisi del contenuto, diretta, in questo caso, ai fini dell'indicizzazione delle materie trattate.

Questo lavoro di metadattazione e di descrizione semantica del contenuto delle risorse digitali, che interessa tanto l'informatica umanistica, quanto l'organizzazione della conoscenza, converge poi significativamente con l'introduzione delle tecnologie che vengono attualmente sviluppate per la costruzione del semantic web. Così, è proprio l'evoluzione tecnologica del web che permette di recuperare quell'aspetto parzialmente negletto dell'informatica umanistica che la sua iniziale costruzione aveva momentaneamente e quasi paradossalmente contribuito a oscurare. Programmaticamente, infatti, il web semantico si pone l'obiettivo dell'elaborazione dell'informazione veicolata dalle risorse digitali che attraverso il web possono essere attinte e visualizzate: «l'elaborazione dei contenuti semantici sta al cuore della visione del *Semantic Web*»³⁶. È quindi naturale che questo sia il terreno di convergenza delle più recenti tendenze della ricerca nel campo dell'informatica umanistica e nelle aree più vicine. La sfida posta dall'evoluzione tecnologica riporta l'informatica umanistica alle sue fondamentali motivazioni originarie.

International ISKO Conference, Rome, 23-26 February 2010 (Advances in Knowledge Organization, 12), Würzburg, Ergon-Verlag, 2010, p. 193.

³⁶ N. Calzolari, "Language resources in the semantic Web vision". In C. Zong (ed), *International Conference on Natural Language Processing and Knowledge Engineering: Proceedings*, Beijing, IEEE Press, 2003, p. 16.

9. Possiamo così concludere queste considerazioni accennando ad un esempio significativo di questa convergenza e ad alcune iniziative di ricerca specifiche nel campo dell'informatica umanistica, il cui orientamento si rivolge manifestamente nella stessa direzione. Il primo esempio riguarda lo sviluppo delle cosiddette *semantic digital libraries*³⁷. L'idea nasce dall'esigenza di sviluppare ed estendere i sistemi di gestione delle biblioteche digitali (DLMS) integrando tecnologie proprie del web semantico (3.0) e del *social networking* (2.0)³⁸. La nuova concezione delle "biblioteche digitali semantiche" si fonda sul presupposto che «le tecnologie semantiche» e dei *social networks* «siano in grado di offrire soluzioni più efficienti per la costruzione di procedure affidabili e facili da usare per accedere ai contenuti e ai metadati»³⁹. In questa prospettiva, le biblioteche digitali sono destinate ad evolversi in «sistemi complessi che integrano servizi avanzati di biblioteca digitale» e nuovi servizi di «sostegno alle comunità di pratica e di ricerca». Per ottenere questo risultato, «in una biblioteca digitale, l'organizzazione della conoscenza entra in gioco in forme diverse e strettamente interrelate» e, più specificamente, come gestione dell'«informazione su termini e concetti e sulle loro relazioni», oppure come «organizzazione di informazione ontologica e lessicale»⁴⁰.

³⁷ Cfr. *Semantic Digital Libraries: Bringing Libraries to Web 3.0*, <http://semdl.info/> (consultato il 7 aprile 2012).

³⁸ Ivi, <http://semdl.info/tutorials/6/abstract> (consultato il 7 aprile 2012).

³⁹ S. R. Kruk et al., *Architecture of Semantic Digital Libraries*, Digital Enterprise Research Institute (DERI), National University of Ireland, Galway, Work Package 4, 30 June 2008, p. 1.

⁴⁰ D. Soergel, "Digital Libraries and Knowledge Organization". In S. R. Kruk and B. McDaniel (ed.), *Semantic Digital Libraries*, Berlin, Springer, 2009, p. 9 e 21.

Il progetto delle biblioteche digitali semantiche si propone come una forma di risposta alla 'nuova preoccupazione' provocata dal web e al timore di perdere quell'«ordine»⁴¹ proprio dei sistemi bibliotecari tradizionali, che lo spazio informativo offerto dal web non riuscirebbe a ricostituire. Si può tuttavia ribattere, a questo proposito, che le tecnologie del web semantico, ovvero di quello che è stato chiamato il *Web 3.0*⁴², offrono «tecniche di ricerca efficienti nel nuovo e interconnesso spazio informativo» delle risorse accessibili in rete. Infatti, lo spazio informativo del *Web* non è uno spazio completamente aperto e privo di struttura, ma possiede una propria forma di organizzazione costituita dai *link* che collegano risorse diverse, o parti diverse di risorse, o documenti digitali. Inoltre, il ricorso alle ontologie, la tecnologia semantica dotata di maggiore potenza inferenziale, produce nuove forme di organizzazione dell'informazione che non si riducono ad «una mera specificazione di schemi di metadati» già definiti e consolidati nella pratica di catalogazione e descrizione bibliografica tradizionale, ma permettono ai metadati «di diventare più aperti, destrutturati e, ciò che più importa, altamente interconnessi». Infatti,

⁴¹ M. Deegan, K. Sutherland, *Transferred Illusions*, Arnham ; Burlington, Ashgate, 2009, pp. 149 e 173.

⁴² «'Semantic Web' è talvolta usato come sinonimo di 'Web 3.0', anche se la definizione di ciascun termine varia», "Semantic Web". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, http://en.wikipedia.org/wiki/Semantic_Web (consultato il 5 aprile 2012); cfr., p. es., J. Markoff, "Entrepreneurs See a Web Guided by Common Sense". In *The New York Times*, 12 novembre 2006, <http://turing.cs.washington.edu/NYT-KnowItAll.htm> (consultato il 9 aprile 2012): «Web 3.0 – o 'Web semantico', dall'idea di aggiungere significato».

le ontologie non introducono vincoli sui metadati, ma permettono piuttosto di ragionare su concetti interconnessi, in quanto assegnano un significato agli oggetti informativi e alle loro relazioni. Quindi, nelle biblioteche digitali semantiche, i metadati diventano una rete di concetti interconnessi e dotati di senso.⁴³

Le biblioteche digitali semantiche intendono inoltre avvalersi dell'integrazione delle tecnologie del web semantico con le forme di indicizzazione e classificazione praticate da comunità di utenti nel cosiddetto *Web 2.0*⁴⁴ e si presentano anche come *social semantic digital libraries*⁴⁵. Così, ad esempio, le *folksonomies*, ossia la «pratica di categorizzazione collaborativa mediante parole chiave scelte liberamente», vengono valutate come «un meccanismo leggero (*lightweight*) e facilmente accessibile per la creazione di ontologie e metadati», tale da permettere l'affiorare di una «semantica emergente (*emergent semantics*)» e rendere concreta «l'aspettativa che le interazioni individuali di un gran numero di agenti razionali producano effetti globali valutabili come una semantica»; in questo modo, «le ontologie diventerebbero un effetto emergente del

⁴³ S. R. Kruk et al., "Architecture of Semantic Digital Libraries". In S. R. Kruk., McDaniel (ed.), *Semantic Digital Libraries*, cit., pp. 78-79.

⁴⁴ «Un sito Web 2.0 permette agli utenti di interagire e collaborare tra loro, dialogando in un *social medium* come produttori-consumatori (*prosumer*) di contenuti generati dagli utenti in una comunità virtuale» ("Web 2.0", in *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, http://en.wikipedia.org/wiki/Web_2.0; (consultato il 9 aprile 2012).

⁴⁵ Cfr. S. R. Kruk et al., *The anatomy of a Social Semantic Digital Library*, http://www.contexta.cl/mediawiki/images/7/72/The_anatomy_of_a_Social_Semantic_Digital_Library.pdf (consultato il 9 aprile 2012).

sistema, anziché costituire una convenzione fissa e limitata della maggioranza». ⁴⁶

Di particolare interesse, a questo proposito, sono le prime implementazioni di sistemi di *social tagging* che permettono di produrre annotazioni nella forma di *linked data*⁴⁷. Si tratta di ontologie cosiddette "leggere" (*lightweight*) sviluppate per l'annotazione, «ontologie di annotazione (*tag ontologies*)» che «permettono di ottenere rappresentazioni interpretabili automaticamente (*machine-readable*) capaci di essere condivise da diversi sistemi di annotazione collaborativa (*social tagging*)»⁴⁸. In questo modo, si realizza l'integrazione tra le tecnologie del web 2.0, il web orientato all'utente, e il web 3.0, il web orientato alle macchine e all'elaborazione automatica dell'informazione. Il risultato più significativo di questa convergenza tra l'inserimento di metadati semantici aggiunti dagli utenti e la loro acquisizione ed elaborazione in un sistema integrato di organizzazione della conoscenza è appunto costituito dalla produzione di *Open Linked Data*, un processo attualmente in corso, di natura quasi evolutiva, paragonabile alla crescita e all'espansione

⁴⁶ P. Mika, "Ontologies Are Us: A unified model of social networks and semantics". In *Journal of Web Semantics*, 5, 2007(1), p. 5-7.

⁴⁷ Il termine «descrive un metodo di pubblicazione di dati strutturati in modo che possano essere interconnessi tra loro e poter essere meglio utilizzati», "Linked data". In *Wikipedia: The Free Encyclopedia*, Wikimedia Foundation, http://en.wikipedia.org/wiki/Linked_data (consultato il 10 aprile 2012).

⁴⁸ H. L. Kim et al., "The state of the art in tag ontologies: A semantic model for tagging and folksonomies". In *DCMI '08: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*, Dublin Core Metadata Initiative, 2008, <http://dc2008.de/wp-content/uploads/2008/09/kim-scerri-breslin-decker-kim.pdf> (consultato il 10 aprile 2012).

iniziale del web, il cui sviluppo assume forme del tutto analoghe a un fenomeno biologico di diffusione virale.⁴⁹

10. Questi recenti sviluppi delle tecnologie informatiche favoriscono il ritorno dell'informatica umanistica alla sua ispirazione originaria: l'«attenzione» che in una successiva fase di sviluppo si era rivolta principalmente «alla rappresentazione delle fonti primarie» torna a rivolgersi «alla costruzione degli strumenti» per l'elaborazione del loro contenuto informativo⁵⁰. Anche per le discipline umanistiche «il *Web* semantico» rappresenta «il futuro» e l'informatica umanistica è chiamata a fornire «rappresentazioni formali del lascito documentario dell'umanità (*human record*)», adatte cioè ad essere elaborate automaticamente. Infatti, come ancora ricorda John Unsworth,

queste rappresentazioni – ontologie, schemi, rappresentazioni della conoscenza, comunque le si voglia chiamare – dovrebbero essere prodotte da persone formate negli studi umanistici. E la disciplina che li produce richiede una formazione umanistica, unita ad una conoscenza di elementi di matematica, logica, ingegneria e informatica. [...] C'è una grande quantità di lavoro da fare – e indubbiamente non tutto di natura tecnica. Nella costruzione di questa grande mappa del

⁴⁹ Cfr. il *Linking Open Data (LOD) Project Cloud Diagram*, <http://richard.cyaniak.de/2007/10/lod/> (consultato il 9 aprile 2012); per il *Linking Open Data community project*, cfr. http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description (consultato il 9 aprile 2012). Nelle intenzioni dei promotori dell'iniziativa, «i *Linked Data* si propongono di usare il Web per collegare dati in relazione tra loro», *Linked Data: Connect Distributed Data across the Web*, <http://linkeddata.org> (consultato il 10 aprile 2012).

⁵⁰ J. Unsworth, *Forms of Attention*, cit.

sapere una grande parte sarà costituita da lavoro collaborativo (social work), creazione del consenso, compromesso. Ma anche quest'attività dovrà essere affidata a persone che sappiano come il consenso possa essere raggiunto ed espresso in un *medium* di natura computazionale.⁵¹

Possiamo ora, in conclusione, ritornare al nostro esempio iniziale che ci invitava a riflettere sulla natura computazionale delle edizioni digitali, per cercare di illustrare le sfide che attendono l'informatica umanistica nella fase presente, contrassegnata dal diffondersi delle attuali tecnologie semantiche. Dopo lo sviluppo delle tecnologie per la *rappresentazione* del testo con l'impiego di marcatori inseriti all'interno della stringa di caratteri che ne costituisce la rappresentazione codificata di base (*embedded markup*), il problema dell'*elaborazione* del contenuto informativo dei dati testuali si presenta in forme nuove.

Attualmente, la rappresentazione critica dei testi si attiene soprattutto al canone impostosi con l'affermarsi e il diffondersi della pratica di codifica testuale fondata sulle norme della Text Encoding Initiative (TEI). Il problema dell'*elaborazione* del contenuto informativo dei dati testuali si pone quindi come il tentativo di mettere in relazione la descrizione delle caratteristiche testuali ottenuta con l'impiego di una forma di *embedded markup*, qual è quella rappresentata dalla codifica TEI, con le possibili forme di indicizzazione o di rappresentazione semantica del contenuto del testo. Il problema non è di facile soluzione, perché le strutture e le proprietà della rappresentazione del testo non sono isomorfe alla struttura e agli elementi della rappresentazione semantica del suo

⁵¹ Id., *What is Humanities Computing*, cit.

contenuto. Non è ancora stato perfezionato e tanto meno generalmente accettato un modello funzionale adeguato, che permetta una mappatura idonea e computabile tra gli elementi delle due strutture⁵². Si possono tuttavia richiamare due recenti proposte che possono fornire indicazioni utili alla soluzione del problema e che rappresentano due esempi significativi del rinato interesse per lo sviluppo di strumenti utili all'elaborazione dell'informazione veicolata dai dati testuali.

La prima iniziativa è costituita dal progetto Humanities Research Infrastructure and Tools (HRIT) promosso dal Center for Textual Studies and Digital Humanities della Loyola University di Chicago, un'iniziativa che si propone di sviluppare strumenti e servizi infrastrutturali per la creazione di un ambiente collaborativo funzionale al lavoro di preparazione di edizioni critiche digitali. Tra gli obiettivi dichiarati del progetto spicca quello di «costruire un ambiente *open-source on-line*, efficiente e collaborativo, in cui aggregare, collegare ipertestualmente (*link*) o con riferimenti incrociati, e ancora emendare e condividere testi esaminati criticamente (*vetted*) di fonti documentarie primarie – corredati con apparati critici e con analisi e commenti, in forma di codifica (*markup*), annotazioni, indicizzazione con parole chiave, collegamenti ipertestuali (*linking*), e così via»⁵³. Come si vede, il progetto mira sia alla rappresentazione dei testi, nella forma di edizioni critiche digitali codificate, sia all'elaborazione e all'analisi del loro contenuto, sviluppando

⁵² Per una discussione più approfondita del problema, cfr. D. Buzzetti, *Oltre il rappresentare*, in *La macchina nel tempo*, cit., pp. 39-62.

⁵³ HRIT Wiki, <https://sites.google.com/a/ctsdh.luc.edu/hrit-intranet/> (consultato il 10 aprile 2012).

strumenti funzionali idonei, compatibili e integrabili con tecnologie semantiche e collaborative.

Secondo Desmond Schmidt, «uno degli obiettivi primari del progetto HRIT», al quale partecipa, «è quello di modificare la tecnologia del *markup* in modo da permettere la collaborazione mediante la condivisione e il riuso dei contenuti». ⁵⁴ La soluzione proposta consiste in una rappresentazione esterna (*standoff*) ed espressa come una stringa di caratteri, di *proprietà* assegnate a pericopi (*ranges*) di testo. Tale soluzione permette di rappresentare, in un unico formato, sia proprietà testuali o dell'*espressione* del testo, sia proprietà interpretative o del *contenuto* del testo. Le variazioni di entrambi gli insiemi di proprietà possono essere rappresentate mediante «un grafo orientato, con un nodo iniziale e un nodo finale» – il formato Multi-Version Document, o MVD⁵⁵ – e ci si può chiedere se ci siano vincoli reciproci che connettono le variazioni di entrambi i grafi, per riuscire a mettere in relazione la rappresentazione delle proprietà testuali, solitamente espressa con forme di *embedded markup*, con la rappresentazione delle proprietà interpretative del testo, solitamente espressa con forme di *standoff markup*. Lo stesso Desmond Schmidt riconosce che la proposta di rappresentare le caratteristiche testuali con proprietà esterne (*standoff properties*) alla stringa di caratteri in cui consistono i dati testuali «non è un'idea originale», ma si richiama «in misura maggiore o minore» ad altre implementazioni come, tra l'altro, «i sistemi di annotazione

⁵⁴ D. Schmidt, "Standoff Properties". In *HRIT: Facilitating interoperability by overcoming overlap*, abstract presentato al DH2012 (in corso di stampa).

⁵⁵ Id., "The Inadequacy of Embedded Markup for Cultural Heritage Texts". In *Literary and Linguistic Computing*. 25, 2010(3), p. 350.

semantica» e il cosiddetto «modello della stringa ampliata (*extended string*)»⁵⁶ proposto da Manfred Thaller.⁵⁷ Ed è questa, in effetti, la seconda linea di ricerca alla quale si intendeva fare riferimento per segnalare, con le parole di Unsworth, nuove «forme di attenzione» allo sviluppo di «strumenti» per l'elaborazione del contenuto informativo dei testi e delle altre risorse digitali accessibili in rete.

La proposta di Manfred Thaller fa riferimento ad un modello per la rappresentazione dinamica dell'informazione trasmessa da un «testo storico», ricavato dall'«*equazione infologica* di Langefors». Nell'implementazione del modello, i «segni materiali (*tokens*) che veicolano l'informazione», sono tenuti «nettamente e inequivocabilmente separati» dal «significato» che viene loro assegnato, sia esso un'istruzione per la loro «resa» grafica, oppure «un descrittore semantico su cui possa operare qualche sistema di recupero dell'informazione (*information retrieval system*)». I diversi tipi di informazione che vengono associati ai segni materiali che la veicolano, ovvero ai dati che la trasmettono, sono rappresentati in uno «spazio interpretativo a *n* dimensioni», che permette di considerare il «testo» come una «stringa di caratteri diversi, ognuno dei quali può essere descritto con un numero arbitrario di proprietà»⁵⁸. Nel quadro del progetto PLANETS, dedicato alla conservazione

⁵⁶ HRIT (*Humanities Research Infrastructure and Tools*), *Standoff Properties*, <http://dhtestbed.ctsdh.luc.edu/hritinfrastructure/index.php/standoff-properties> (consultato il 12 aprile 2012).

⁵⁷ Cfr. M. Thaller, "Strings, Texts and Meaning". In *Digital Humanities 2006: Conference Abstracts*, Paris, CATI - Université Paris-Sorbonne, 2006, p. 212-14.

⁵⁸ Id., *What Is a Text within the Digital Humanities, or Some of Them, at Least?*, abstract presentato al DH2012 (in corso di stampa).

D. Buzzetti, *Che cos'è, oggi, l'informatica umanistica?*

digitale⁵⁹, questo modello astratto è stato concretamente applicato allo «sviluppo di strumenti»⁶⁰ per confrontare il contenuto informativo di rappresentazioni diverse, secondo diversi formati, di uno stesso oggetto culturale⁶¹. Anche in questo caso, come si vede, la rappresentazione digitale dell'oggetto e l'elaborazione del suo contenuto informativo vengono chiaramente distinte e considerate degne di pari attenzione.

Lo sviluppo dell'informatica umanistica sembra decisamente avviarsi, nella fase attuale, in questa direzione.

⁵⁹ Cfr. <http://www.planets-project.eu/about/> (consultato il 12 aprile 2012).

⁶⁰ M. Thaller, *What Is a Text*, cit.

⁶¹ Cfr. Id., *The eXtensible Characterisation Languages – XCL*, Hamburg, Kovač, 2009.

Il testo e il computer

Da una digital library del latino tardo ad un corpus globale

Maurizio Lana

Origine del progetto

digilibLT ("digital library of late Latin Texts") è una digital library in corso di realizzazione destinata a raccogliere i testi latini tardi, cioè quei testi che si collocano nel periodo tra la fine del I-inizio del II secolo e il VI secolo d.C. Il progetto è finanziato dalla Regione Piemonte in quanto – insieme ad altri - è risultato vincitore del bando per la ricerca nelle scienze umane e sociali pubblicato nel 2008. Nasce da un'idea di Raffaella Tabacco e da un progetto di Maurizio Lana tenendo conto di ciò che oggi si chiede ad una digital library aperta, fruibile, resiliente. Il progetto è guidato dall'Università del Piemonte Orientale, Dipartimento di Studi Umanistici, e vede la collaborazione dell'Università di Torino, Dipartimento di Filologia, Linguistica, Tradizione Classica, anche a causa di stretti legami di formazione, di studio e di ricerca tra i latinisti delle due istituzioni.

L'idea originaria fu di completare la raccolta di testi latini su CDROM che gli studiosi di cose classiche conoscono come PHI cdrom 5.3 (PHI sta per Packard Humanities Institute). Si tratta di una raccolta di tutti i testi latini dalle origini fino a fine I-inizio II sec. d.C.(più altri, sparsi), provenienti da edizioni critiche e privi di apparato critico. Essa presenta una caratteristica rilevante, sia per i tempi in cui fu creata, sia ancora oggi: la codifica dei testi.

La codifica dei testi del PHI e i suoi limiti

I testi del PHI sono in un formato ibrido, un TXT che contiene una codifica proprietaria nota come Beta code, sviluppata dal PHI e poi usata anche per il cdrom del TLG (*Thesaurus Linguae Graecae*), che contiene tutti i testi greci dall'epoca arcaica all'epoca bizantina. In tale codifica sono descritti i livelli di struttura del testo (autore, titolo, numerazione degli elementi strutturali come libro, capitolo, paragrafo, riga), i fenomeni filologici, i caratteri greci ed ebraici e altri simboli (si ricordi, per comprendere, che la codifica beta nacque quando il codice ASCII a 256bit era tutto ciò che si poteva usare per rappresentare i caratteri greci in formato TXT). Pur essendo nota con un certo livello di dettaglio, essa non fu mai *intenzionalmente* descritta pubblicamente in tutti i suoi aspetti. L'elemento più caratteristico è che si tratta di un codifica compatta, implicita. Infatti, allo scopo di non sprecare nemmeno 1 bit, i vari livelli strutturali e i loro valori numerici sono descritti utilizzando *blocchi di bit all'interno di alcuni byte* collocati all'inizio dell'unità testuale da descrivere. Il risultato è che nel complesso delle opere si risparmiano, sul totale del cdrom che contiene i testi, un po' di byte ma al prezzo di rendere

la descrizione della struttura del testo assolutamente incomprendibile per chi non possieda le informazioni necessarie e magari anche uno strumento software adatto. Chi aprisse con un editor il file di testo che contiene la "vita di agricola" scritta da Tacito troverebbe questa stringa iniziale:

```
ï€±³µ±ÿi°°±ÿi,ÁçÿifÔáäÿ-ôÿ@@@{1DE VITA €@@@@IVLII  
AGRICOLAE €@@@@LIBER}
```

nei cui caratteri "strani" sono contenute queste informazioni (rese esplicite con il programma Diogenes di P. Heslin):

```
chapter t, section 1, line 1,  
DE VITA  
IVLII AGRICOLAE  
LIBER
```

La codifica TEI

Chi oggi pensi al formato e alla codifica dei testi di una digital library pensa a standard pubblici, condivisi, e a codifiche esplicite, resilienti, che garantiscano di durare nel tempo. Pur con qualche cautela critica si pensa a TEI, dove il problema è probabilmente di non perdere il bandolo nei meandri delle infinite finzze rese possibili nella descrizione del testo attraverso tale codifica¹. Quindi i

¹ *Guidelines for Electronic Text Encoding and Interchange*, 2 voll., Oxford, Providence, Charlottesville, Nancy, TEI Consortium, 2008. In italiano si veda l'edizione ridotta delle linee guida, pubblicata a cura di F. Ciotti: Lou Burnard, C. M. Sperberg-McQueen, *Il manuale TEI Lite : introduzione alla codifica elettronica dei testi letterari*, Milano, Sylvestre Bonnard, 2005.

testi latini tardi raccolti in digilibLT sono marcati secondo lo standard TEI allo scopo di descrivere/gestire:

- la struttura dei testi (libro, capitolo, paragrafo, riga, in termini generali; poi ogni opera può avere sue specificità da questo punto di vista);
- i fenomeni filologici principali: integrazioni, espunzioni, lacune, *cruces*;
- la presenza di immagini (tra i testi tardi ve ne sono molti di argomenti scientifico, ed essi, per esempio, contengono figure geometriche);
- la presenza di simboli (per lo più unità di misura) i cui glifi non sono tutti disponibili nei font correntemente in uso (Arial, Times, eccetera) o nei font speciali per l'antichità (Cardo, su tutti); vale a dire che mentre i glifi di *pondo*, *dextans*, *dodrans*, e altri, si possono rappresentare o comporre con i normali font, al contrario unità di misura come il *dupondius*, o il *denarius*, e altre ancora, sono prive di glifo tra i font normalmente diffusi; e TEI prevede la possibilità che un glifo sia rimpiazzato dal nome del simbolo che rappresenta.

Se digilibLT è la "seconda parte", il completamento, della raccolta dei testi latini digitalizzati di cui il PHI cdrom costituisce la prima, chiaramente si pone un problema di omogeneità e continuità nella codifica. A tale scopo è stato previsto di distribuire i testi digilibLT anche codificati in formato Beta, così da garantire la compatibilità in basso per tutti gli studiosi già abituati a specifici strumenti software di gestione della digital library PHI 5.3. Non è da escludere che si possa anche arrivare all'altra soluzione simmetrica a questa: marcare in standard TEI nel modo indicato sopra tutta la

raccolta PHI 5.3. La cosa comporta, ovviamente, una certa complessità di questioni legali da dipanare.

Che cosa il progetto offre agli studiosi

L'attività principale, svolta con la preziosa, competente e insostituibile collaborazione di dottori di ricerca, dottorandi, laureati magistrali e laureandi², è quella dell'edizione digitale. Nelle due sedi del progetto i testi vengono digitalizzati per mezzo di scanner Atiz Bookdrive, poi le immagini vengono impacchettate in PDF, e passate al software di riconoscimento caratteri. L'esito dell'OCR viene riscontrato e corretto sugli originali in due differenti fasi. La codifica TEI in parte viene inserita o predisposta durante la correzione (eseguita in Word o in text editor a seconda delle preferenze delle persone) e in parte in una fase di lavoro specifica con Oxygen, un editor XML specializzato. Si tratta a tutti gli effetti di un lavoro di edizione, che comporta valutazioni e scelte scientifiche nel dominio latinistico, rese poi operative attraverso gli strumenti digitali.³

La digital library (<http://digiliblt.unipmn.it>, *Figura 1*) offre agli utenti del sito una varietà di servizi e informazioni che si possono raggruppare in 4 categorie principali:

- accesso ai testi e bibliografia selezionata
- ricerche sui testi

² Simona Musso, Alice Borgna, Manuela Ferroni, Laura Mosca, Nadia Rosso, e altri.

³ Solo il differente livello di acquisizione e inculturazione della stampa rispetto al digitale può indurre a ritenere che nelle attività di preparazione editoriale di un libro non vi sia lo stesso livello di complessità tecnico/tecnologica presente nella preparazione di un testo per l'edizione digitale.

- risorse sul tema del latino tardoantico
- risorse informatiche

Figura 1.

The screenshot displays the digilibLT website, titled "Biblioteca digitale di testi latini tardoantichi". The header includes the project name, a logo, and the logo of the Regione Piemonte. A navigation bar contains links for Home, the project, Tardoantico nel web, Auto, Contatti, and Feedback. A search bar is prominently featured with a "CERCA" button. Below the search bar, there are sections for "Opere disponibili per la ricerca" and "Ricerca bibliografica". A map of Italy highlights the regions of Novara, Vercelli, and Alessandria. The main content area is divided into several sections: "Letture e download di testi" with options for Opere, Autori, and Bibliografia; "Per ordine cronologico" (II, III, IV, V, VI, VII) and "Datazione incerta"; "Per ordine alfabetico" (A-Z, O-R, S-T, V); "Ricerca" with links to Opere sul tardoantico, Canone autori tardoantichi, Download di font e software, and Scarica i testi in formato PHF/TLG; and "Proponente aggiuntivo" with a circular logo. The news section includes "Due nuovi testi disponibili" (27/2/2012) about Vergilius orator and Opusculum, "Nuove accessioni su DigilibLT" (1/2/2012) about Gratianum actio and Tardae passiones, and "Il corpus di Sesto Aurelio Vittore è online su DigilibLT" (13/1/2012) about the Liber de Caesaribus and Corpus tripartitum.

Accesso ai testi e bibliografia

L'accesso ai testi può avvenire attraverso una selezione per titolo dell'opera oppure per autore e titolo dell'opera. Individuata l'opera, al lettore vengono presentate varie informazioni e scelte:

Figura 2.

Res gestae Alexandri Macedonis
Tullius Valerius (Tullius Valerius Alexander Polemius) [Scheda Autore](#)

Post 337

[LEGGI IL TESTO](#)

Download possibili

[TXT](#) [TEI](#) [BETA](#) [PDF](#) [E-PUB](#) [SCHEMA CATALOGRAFICA](#)

Edizione di riferimento:

Iulii Valerii Res Gestae Alexandri Macedonis translatae ex Aesopo Graeco, adhibitis schedis Roberti Calderan edidit Michaela Rosellini, editio correctior cum addendis, Monachii et Lipsiae 2004 (Bibliotheca Scriptorum Graecorum et Romanorum Teubneriana).

Le *res gestae Alexandri Macedonis* di Giulio Valerio sono la traduzione latina del cosiddetto *Romanzo di Alessandro*, falsamente attribuito dalla tradizione a Callistene. L'opera fu scoperta da Angelo Mai all'inizio dell'800 in un codice Ambrosiano (X-IX sec.), che contiene anche l'*Itinerarium Alexandri* (v. scheda apposta), e da lui edita per la prima volta nel 1817. Il testo, lacunoso nel codice Ambrosiano, è conservato in modo più completo

Acquisizione immagini: Alice Borgna
Preparazione testo: Raffaella Tabacco
Codifica TEI: David Paniagua
Edizione digitale a cura del gruppo digilibLT - Università degli Studi del Piemonte Orientale

[Torna all'elenco](#)

Bibliografia

1. André, J. Échos poétiques d'un brahmane
2. André, J; Filozat, J L'Inde vue de Rome. Textes latins de l'Antiquité relatifs à l'Inde
3. Arduini, Paolo et alii **Studi offerti ad Alessandro Perutelli, vol. 1**
4. Ausfeld, A. Der griechische Alexanderroman
5. Ausfeld, A. Zur Topographie von Alexandria und Pseudokallisthenes
6. Axelsson, S. Zum Alexanderroman des Julius Valerius. Textkritische Beiträge
7. Berengo Giovanni Le **imprese di Alessandro Macedone scritte in greca favella da Esopo e recate in latina da Giulio Valerio, scoperte dall'Em. Cardinale**

- l'accesso alla lettura full text dell'opera;
- la possibilità di scaricare l'opera in uno di vari formati disponibili;
- una scheda catalografica descrittiva molto concisa;
- una scheda critica originale firmata da uno dei latinisti del gruppo di ricerca;
- informazioni sul processo di lavorazione del testo con i nomi delle persone responsabili delle fasi principali
- bibliografia selezionata.

In che cosa consista la "lettura full text dell'opera" (che è accompagnata sulla sinistra da una mappa del documento che facilita al lettore lo spostamento nel punto desiderato) è abbastanza chiaro, mentre può essere utile soffermarsi sugli altri elementi. Il download del testo è libero, non comporta costi, solo la necessità di registrarsi sul sito perché ciò permette al gruppo di ricerca di documentare l'attività di studio innescata. Il download è gratuito perché si adotta una politica di accesso aperto⁴ combinata con licenze Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo 3.0⁵. Il download è legale perché il testo stabilito di un'edizione critica, secondo la normativa italiana sul diritto d'autore, è proprietà intellettuale dell'autore antico, dalla cui morte sono trascorsi più di 70 anni; mentre appartiene al curatore l'apparato critico.

Per "scaricare i testi" sono previsti vari formati, alcuni già attivi e altri attivati a breve; ognuno di essi risponde ad una differente tipo di utilizzo: il formato TEI (cioè TXT con tag XML) permette di lavorare con strumenti software specificamente orientati a tale codifica e costituisce anche la base da cui gli altri formati vengono

⁴ Sul tema dell'open access si veda per esempio J.-C. Guédon, *Open access: contro gli oligopoli nel sapere*, Pisa, ETS, 2009. Molte università italiane hanno aderito ai principi dell'open access esposti nella dichiarazione di Berlino (<http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>) a partire dalla dichiarazione di Messina, il cui testo è disponibile all'indirizzo <http://www.aepic.it/conf/Messina041/index981f.html>. Ai primi firmatari altri se ne sono aggiunti (<http://www.cruil.it/HomePage.aspx?ref=894>).

⁵ Le licenze Creative Commons nascono da un'idea del giurista statunitense L. Lessig che ha scritto quattro libri principali sul tema: nel 2001 *The Future of Ideas*, nel 2004 *Free Culture: The Nature and Future of Creativity*, nel 2006 *Code: Version 2.0* e nel 2008 *Remix: Making Art and Commerce Thrive in the Hybrid Economy*.

ricavati. Il formato TXT è pensato per chi vuole poter lavorare sui testi con programmi per analisi statistiche multivariate, con programmi di text retrieval, e simili. Il formato Beta code permetterà di studiare i testi utilizzando uno dei programmi ancora oggi in uso per lo studio dei testi del PHI cdrom (Musaios, Diogenes, e simili); in formato Beta saranno disponibili anche le immagini ISO – progressivamente aggiornate nel tempo, seguendo l'avanzamento del progetto - che permetteranno di produrre un cdrom contenente la base di dati testuale in formato Beta. Il formato PDF renderà agevole la stampa impaginata, pensata soprattutto per chi ama avere i testi stampati su carta per poterli annotare. Il formato e-pub renderà invece possibile portare con sé e leggere i testi (anche, ma non solo) su dispositivi mobili come tablet, creando così un nuovo tipo di biblioteca personale. I formati TXT, Beta, PDF, e-pub sono tutti ricavati a partire dal testo base marcato in XML-TEI, per mezzo di fogli di trasformazione. Ciò ha lo scopo di evitare il rischio di un disallineamento delle varie versioni dei file.

La "scheda catalografica", ancora non completamente sviluppata, permetterà l'identificazione delle opere da parte di /all'interno di digital libraries più ampie: Internet Culturale ed Europea in primis; il processo formale e tecnico per l'inserimento della digital library digilibLT all'interno di Internet Culturale è stato avviato.

La "scheda critica" descrive l'opera dal punto di vista filologico-storico-letterario così da aiutare tutti coloro – studiosi, studenti, appassionati, curiosi – che si avvicinano alla conoscenza di questi testi senza una formazione specifica. Pur nella sua concisione è comunque uno scritto critico originale ed è dunque firmata.

Uno "spazio di credits cita le persone responsabili delle varie fasi principali della lavorazione del testo". Il senso è che è necessario

essere attenti anche al lavoro intellettuale di chi ha reso possibile la pubblicazione del testo.

Infine uno spazio di "bibliografia selezionata" mostra gli studi più importanti esistenti sul testo antico in esame, sul modello di quanto sviluppato da A. Balbo ed E. Malaspina per i siti latinistici Tulliana⁶ e Senecana⁷ da loro curati. Si tratta di citazioni bibliografiche utili per approfondimenti critici, non di link al full text, ovviamente. Un possibile sviluppo di questa sezione sarà il link al record corrispondente su SBN così da poter individuare in quale biblioteca si trova quel certo studio critico.

Ricerche online sui testi

Il sito rende possibile effettuare ricerche online, con la consueta duplice interfaccia "semplice" e "avanzata", senza dimenticare la ricerca all'interno dei titoli di bibliografia. Le valutazioni sono complesse e sfaccettate, sul tema delle ricerche online all'interno di basi di dati. Da una parte le ricerche online non possono raggiungere la potenza e la flessibilità offerte da programmi anche semplici di text retrieval (MonoConcPro, o KwicKwic, ne sono ottimi esempi); dall'altra possono essere effettuate al volo, senza compiere operazioni informatiche su file e programmi. La ricerca semplice cerca in tutti i testi disponibili sul sito la/e parola/e indicate dell'utente; mentre la ricerca avanzata offre i tipici criteri di selezione:

⁶ <http://www.tulliana.eu>

⁷ <http://www.senecana.it>

Figura 3.

Ricerca all'interno dei testi

 tutte (AND) almeno una di (OR) queste parole
Racchiudi le parole tra apici (") per cercare sequenze esatte

Parole da escludere:

Prossimità parola(e)
Sezione testuale
 tutto il testo titoli verso passo in greco

Criteri di selezione delle opere
Titolo
 tutte almeno una di queste parole
Autore
 tutte almeno una di queste parole
Tipologia testuale
Datazione II III IV V VI VII Datazione incerta

criteri di selezione delle parole da cercare: collegamento in AND o in OR, esclusione di parole, indicazione delle aree (titoli, body text, greco) in cui effettuare la ricerca; e criteri di selezione delle opere in cui effettuare la ricerca: parole del titolo, autore, tipologia testuale e epoca in cui si colloca l'opera. Tipologia testuale e epoca sono criteri di selezione perché ogni opera quando viene caricata nel sistema è accompagnata da una scheda che la descrive e le associa – appunto – informazioni sulla tipologia testuale e sull'epoca. Questi stessi due criteri possono essere usati per raffinare la visualizzazione degli esiti di una ricerca:

Figura 4.

The screenshot displays a search interface for the word "imperator". At the top, it indicates "Ricerca eseguita: **imperator** in tutti i testi" and "Risultati: 95 occorrenze in 19 testi". There are links for "Modifica ricerca" and "Nuova ricerca". Below this, the "Ordinamento" is set to "n. occ." with a "Vai!" button, and the page number is "Pagina: 1".

The main content area is divided into two columns. The left column, titled "Tipologia testuale [-]", lists various categories with their respective counts: Astronomia (1), Biografia (2), Botanica (1), Epistolografia (2), Fisiognomia (2), Geografia (2), Grammatica (1), Ingegneria (1), Medicina (1), Mineralogia (1), Mitografia (3), Polemografia (2), Retorica (1), Romanzo (3), Scienza (1), Storiografia (8), Tecnologia (2), Traduzione (1), and Zoologia (1). Below this is a "Data" section with counts for Saec. II (2), Saec. III (2), Saec. IV (13), Saec. V (4), and Saec. VI (3).

The right column displays three search results, numbered 1, 2, and 3. Each result includes the author, title, typology, date, frequency, and concordance information, along with a link to a detailed page.

- 1** Autore: Aurelius Victor Ps. (Corpus triperitum uel Aurelianum)
Titolo: Epitome de Caesaribus
Tipologia testuale: Biografia | Storiografia
Datazione: fere 400
Occorrenze: 16 [Mostra]
Concordanze: Ordinamento Testo - Contesto sinistro - Contesto destro
[Scheda dettagliata dell'opera](#)
- 2** Autore: Ausonius
Titolo: Gratiarum actio
Datazione: n. fere 310 m. fere 394
Occorrenze: 14 [Mostra]
Concordanze: Ordinamento Testo - Contesto sinistro - Contesto destro
[Scheda dettagliata dell'opera](#)
- 3** Autore: Florus
Titolo: Epitoma de Tito Liuio
Tipologia testuale: Storiografia
Datazione: saec. II
Occorrenze: 13 [Mostra]
Concordanze: Ordinamento Testo - Contesto sinistro - Contesto destro
[Scheda dettagliata dell'opera](#)

Si vede infatti sulla sinistra l'elenco delle tipologie testuali in cui si collocano le opere che forniscono esiti per la ricerca della parola "imperator". Cliccando su "storiografia" si restringe la visualizzazione degli esiti alle sole opere la cui scheda descrittiva (interna al sito) ne descriva la tipologia come "storiografia" appunto.

Analogamente per i secoli che indicano la datazione. Le selezioni (i raffinamenti) su tipologia e datazione si integrano: cliccando su "storiografia" e su "saec. IV" si selezionano le opere di argomento storico del IV secolo.

Come si può osservare, gli esiti vengono mostrati suddivisi per opera, non tutti insieme in un'unica grande lista, e si può scegliere di vederli o cliccando su "Mostra", nella riga occorrenze, oppure su uno dei link nella riga "Concordanze", a seconda che si voglia vedere la concordanza KWIC nell'ordine naturale del testo, oppure ordinata

per parola a sinistra o parola a destra della parola cercata (così da mettere in evidenza eventuali cooccorrenze ricorrenti. La scelta di presentare gli esiti della ricerca divisi per opera è strettamente connessa con il fatto che si tratta di esiti di ricerca online: se 3 o 4 utenti contemporaneamente formulassero *queries* che forniscono un numero altissimo di esiti, e gli esiti venissero forniti come concordanza KWIC globale, la cosa potrebbe dar luogo a un forte degrado delle prestazioni del servizio.

Risorse sul tardoantico

Il sito offre link ad altri due tipi di risorse sul tardoantico: il canone degli autori tardo-antichi che saranno via via caricati sul sito (la pubblicazione stessa del canone in questa fase non è ancora completa ed è periodicamente aggiornata) cioè l'elenco dettagliato autori/opere con indicazione dell'edizione critica di riferimento⁸; e studi critici sul tardoantico disponibili in full text, o perché liberi da diritti o perché resi disponibili dai titolari dei diritti.

Risorse informatiche di supporto

- il download di font e software specifico, in particolare il font *Cardo*, che contiene molti glifi relativi a simboli di vario genere usati nei testi classici e tardoantichi, e il programma *CATMA* sviluppato all'Università di Amburgo, che - pur nascendo come

⁸ Il riferimento ideale è il canone del TLG (L. Berkowitz, K. A. Squitier, W. A. Johnson (a cura di), *Thesaurus Linguae Graecae canon of Greek authors and works*, New York, Oxford University Press, 1990). Canone non nel senso di regola a cui conformarsi, ma piuttosto elenco ufficiale dei testi contenuti nella raccolta con l'indicazione delle edizioni a cui si rifanno i testi digitali.

ambiente di marcatura di testi in XML - può essere utilizzato anche per lavorare su testi già marcati in XML (come sono i testi marcati in XML/TEI del progetto digilibLT);

- il download dell'immagine ISO dei file marcati in Beta code: è pensato soprattutto per la compatibilità in basso verso tutti coloro che sono abituati a studiare i testi latini classici utilizzando il cdrom 5.3 del PHI, perché si offre la possibilità di generare un "digilibLT cdrom 1.0" che contiene tutti i testi tardoantichi man mano disponibili del tutto analogo al cdrom del PHI e quindi utilizzabile con gli stessi programmi di text retrieval.

Ampliare gli orizzonti: un corpus globale del latino

Tra i principali corpora di latino digitale pagano e/o cristiano si annoverano oggi (senza voler fare a torto a nessuno per eventuali dimenticanze):

- PHI cdrom 5.3: il latino dalle origini al I sec. d.C.;
- digilibLT (digiliblt.unipmn.it): il latino tardoantico pagano dal I al VI sec. d.C. ;
- Musisque Deoque (www.mqdq.it): il latino poetico pagano e cristiano dalle origini al Rinascimento;
- Bibliotheca Iuris Antiqui(www.bia.net): il latino giuridico;
- Grammatici Latini: il corpus degli scritti dei grammatici latini, dal I sec. a.C. al VI sec. d.C.;
- Archivio della Latinità Italiana del Medioevo (www.uan.it/alim): testi latini medievali di area italiana;
- Corpus Rythmorum Musicum (www.corimu.unisi.it): testi musicali medievali.

Fino ad ora i corpora digitali del latino sono stati sviluppati (o si stanno sviluppando: perché vari di essi sono progetti in evoluzione) in autonomia, e quindi in isolamento reciproco. E' però accaduto che in generale questi progetti pur nella loro separatezza di progettazione e gestione hanno adottato un medesimo approccio molto importante: la codifica dei testi, non la pura digitalizzazione, realizzata secondo le linee guida TEI; con la descrizione di un nucleo di base sostanzialmente comune di fenomeni: la struttura del testo e i fenomeni filologici. In occasione di un recente convegno (Vercelli, 8 Novembre 2011, *Tecnologie moderne per le opere degli antichi*) i responsabili dei progetti di digitalizzazione ricordati hanno convenuto di operare per arrivare a creare un corpus globale del latino, o in altri termini una biblioteca globale del latino. Per una volta si potrebbe impiegare a proposito l'aggettivo virtuale: perché ogni gruppo resta responsabile della propria raccolta, ma accetta di conferire in un repository centrale – per mezzo di operazioni di harvesting – i testi che sono oggetto dei suoi studi. Il repository centrale servirà solo – nell'idea che si sta sviluppando e sulla quale è già stato depositato un progetto per il bando europeo ERC Synergy – per effettuare le ricerche; i passi che interessano verranno aperti in full text dai siti dei progetti. A parte le complesse questioni per il mantenimento dell'allineamento e della coerenza dei due esemplari dei corpora (quello locale e quello centrale), il concetto è chiaro: gli strumenti digitali *permettono* di superare barriere spazio-temporali non solo nella comunicazione personale ma anche nella costruzione degli oggetti di studio, il che prospetta una nuova stagione della *respublica studiorum* con il suo fondamento nei libri (perché non esisterebbero gli studio se non ci fossero i libri a fornire linfa vitale).

Nessuno studioso ad oggi ha mai avuto a disposizione per i suoi studi una biblioteca così ampia come quella in progetto: il puro fatto

di averla a disposizione non basta, per farne buon uso occorrerà immaginare e capire in quali modi essa possa dispiegare il suo potenziale.

Al di là del latino

Il progetto qui sinteticamente descritto al di là della specifica materia costituita dai testi latino tardi, presenta due caratteristiche di sistema che può essere utile evidenziare.

In primo luogo un'attenzione specifica alle condizioni d'uso dei testi da parte degli studiosi. Il progetto è interamente finanziato con fondi pubblici e quindi nella logica dell'Open Access che i suoi risultati sono accessibili pubblicamente a tutti senza costi. Ma accesso aperto non può e non deve significare uso incontrollato dei testi da parte di chiunque, per esempio per fini di lucro. Si è perciò deciso di porre tutto il contenuto del progetto (sito e testi) sotto licenza Creative Commons "Attribuzione, Non commerciale, Condividi allo stesso modo 3.0", per agevolare al massimo il lavoro degli studiosi e la circolazione dei testi. Non sono molte – per così dire – oggi le digital libraries di testi antichi che definiscono esplicitamente le condizioni d'uso dei loro testi in forma giuridicamente chiara e definita (ma in futuro ciò sarà sempre più importante).

In secondo luogo la creazione di un ambiente complesso di lavoro sui testi. Infatti i testi:

- sono disponibili in molteplici formati che rispondono a differenti necessità di lavoro dello studioso (i formati del download ma anche il formato Beta per la compatibilità in basso);

- sono affiancati dalla bibliografia specifica *selezionata*, ed è in studio il modo per rendere possibile dalla scheda dell'opera di bibliografia l'accesso al corrispondente record in SBN, così da poter localizzare il volume che interessa leggere;
- sono tutti introdotti da una sintetica scheda critica che è prodotto originale degli studiosi che collaborano al progetto;
- sono accompagnati da opere critiche di valore sul tardoantico accessibili direttamente in full text;
- possono essere studiati per mezzo degli strumenti di ricerca disponibili sul sito stesso.

A ciò si aggiungono:

- una newsletter settimanale molto concisa che informa delle pubblicazioni di nuove opere nel sito; essa viene inviata alle persone che hanno dato il consenso;
- un piccolo spazio di risorse informatiche utili per i frequentatori del sito;
- uno spazio di interazione con gli studiosi del progetto.

Tutto ciò configura il sito come uno spazio di lavoro a tutti gli effetti, non autosufficiente ma sufficientemente ricco e dotato da risultare accogliente per lo studioso, e per quanto possibile flessibile. Quindi non solo distribuzione di testi in versione mono-formato, ma assistenza e servizi.

1991-2011, un ventennio di evoluzione nella filologia pratica

Paolo Mastandrea

Giusta o sbagliata che sia tale notizia, Wikipedia stabilisce 'ufficialmente' la nascita del Web al 1991, il 6 agosto per la precisione, quando Tim Berners-Lee collocava on-line il primo sito al mondo su Internet¹; e proprio a quell'altezza temporale, sfruttando la suggestività di una cifra che allude a vari precedenti storici o narrativi del passato, il nostro titolo individua *vingt ans après* un memorabile punto di scansione, quasi di partenza per i progressivi sviluppi della nostra cultura. Non sarà forse lo stesso per tutte le discipline accademiche, ma se oggi un filologo, ricercatore nel campo delle letterature antiche e moderne, scorresse con lo sguardo i secoli alle spalle, mai (per esempio fra il 1881 e il 1911, o fra il 1791 e il 1811, eccetera fino a risalire agli stessi incunaboli della stampa): mai, dicevo, saprebbe trovare mutamenti paragonabili per importanza a quelli da noi sperimentati in così breve spazio.

¹ <http://info.cern.ch>.

Al centro esatto dell'arco di tempo – era il 2001, dunque una data cruciale, capace di evocare le più forti emozioni collettive – sta la scomparsa di Giuseppe Gigliozzi, che oggi commemoriamo; e il prematuro ammutolirsi della sua voce appare caso forse più ingiusto e doloroso perché proprio allora andavano compendosi alcuni degli sviluppi concreti da lui previsti, sia riguardo alla costituzione dei grandi *corpora* e archivi elettronici consultabili su CD-ROM o in linea, mediante appositi programmi di interrogazione verbale, sia ai diversi modelli di edizione critica digitalizzata, per cui è oggi possibile rappresentare efficacemente le situazioni testuali più complesse.

Grazie all'impiego dei nuovi strumenti elettronici per la lettura e la scrittura, con l'introduzione delle risorse di rete adatte a specifici scopi di ricerca e controllo bibliografico, si sono realizzati vantaggi non meno fondamentali nella pratica quotidiana degli studiosi, degli studenti, insomma di tutti gli amanti e fruitori della letteratura; talvolta, ben al di là di una piena coscienza degli interessati. D'altro lato sono cadute molte solide certezze riguardo alle qualità del testo e ai poli concettuali entro cui – almeno nella nostra percezione – esso oscilla inevitabilmente: 1) fissità *vs* mobilità, 2) assolutezza *vs* relatività, 3) unitarietà *vs* molteplicità. Prendiamoli uno ad uno, anche se le partizioni sono empiriche e i confini non appaiono così netti.

1. Il testo digitale si presenta nella forma più libera che l'uomo abbia mai conosciuta finora, svincolato da ogni condizionamento esterno; da ogni strumento materiale o elemento fisico di supporto (lo stilo e la penna; i fogli di carta e quant'altro di più pesante duro); da ogni evento sfavorevole che non siano un crash della macchina o un black-out dell'energia elettrica; a differenza del manoscritto, dello

stampato, del dattiloscritto, fin che si trova sullo schermo del nostro PC esso appare divagante e immateriale (e però non astratto) come il pensiero; può incorrere semmai nel rischio della dismisura, della sregolatezza, della provvisorietà; soggetto con minimo sforzo (e nessuna spesa) ad alterazione, si lascia aggiornare correggere rielaborare migliorare, praticamente all'infinito; svapora così l'idea tutta illusoria dell'esistenza stessa di un "testo definitivo": quello che l'intuito geniale di Borges aveva del resto già sospettato di appartenere "soltanto alla religione, o alla stanchezza" dell'autore.

2. Il testo digitale ci fa liquidare d'un tratto ogni superstiziosa riverenza nei confronti del *ne varietur* – comandamento che nasconde puri miti e aspirazioni idealizzate, non solo in tempi pre-gutenberghiani: al punto che i prototipi di documenti politici o prescrittivi diversi – dagli antichi libri divinatori e liturgici, alle raccolte giuridiche, alle carte costituzionali degli stati moderni – tendono a divenire oggetto di una specie di culto, anche civile e profano; più apparente che reale. Nel corso dei secoli (e delle copie, e delle tirature a stampa) proliferano infatti gli avvicendamenti, anche quasi impercettibili sul piano formale: non solo le grafie si rinnovano (*i / j, u / v* in latino, *ho / ò, hai / ài* ecc., in italiano), ma anche le parole si sostituiscono ad altre, cambiano di significato, seguono insomma l'evoluzione dei tempi e le dinamiche delle circostanze sociali, politiche, religiose (censure, eufemismi, metafore, proibizioni di parole straniere, ecc.).

3. Il testo digitale contribuisce indirettamente a far capire meglio in che forme e modi il testo circolasse nell'antichità e nel medioevo europeo. Sino a poco fa, secoli di consuetudine ininterrotta e pressoché esclusiva con i caratteri impressi sulla carta ci avevano

assuefatti a riporre tranquilla fiducia su ciò che riceviamo in lettura per questa via; l'affermarsi di un'editoria commerciale diffusa, soltanto alla fine dell'Ottocento regolata da leggi internazionali sul copyright, portava a considerare sempre 'autorizzato' il libro su cui scorriamo gli occhi; ed ancor oggi, nessuna persona di media cultura dubiterebbe che la stampa delle *Ultime lettere di Jacopo Ortis*, o de *I promessi sposi*, o de *Il fu Mattia Pascal* ripassata sui banchi di scuola, o trovata in casa, o comprata nuova dal libraio, corrisponda alla copia letta già da suo padre e suo nonno, e lo stesso per il suo ex-compagno di banco; da mettere in conto sarà qualche errore tipografico, ma nella sostanza la "verità" del testo unico licenziato dall'autore è fuori discussione. Al massimo, gli studenti più maliziosi e scaltriti potevano notare come alcune forme di antologizzazione dei 'classici' avevano scopi non solo funzionali agli usi scolastici, ma obbedienti a scopi di censura moralistica: ancora dopo la fine degli anni Sessanta, nelle aule dei licei italiani *Furioso* e *Gerusalemme* potevano circolare in edizioni integrali, salvo certe ottave che lasciavano spazio qua e là per compendi in prosa di circospetti annotatori, o brutalmente sostituite da puntini di sospensione; il che stuzzicava l'immaginazione dei ragazzi, avviando i migliori principianti alle loro autonome, utilissime pratiche di ricerca bibliografica e comparazione fra testi, condotte a casa o in biblioteca.

Per capire quali problemi si ponevano nell'era "ante PC", seguendo il procedere di una ricerca nelle sue singole tappe già dagli adempimenti iniziali, è consigliabile procurarsi una copia di *Come si fa una tesi di laurea* di Umberto Eco, fortunato manualletto di 249 pagine, uscito la prima volta per i Tascabili Bompiani nel luglio del 1977. L'autore nutriva una curiosità precoce per la sopraggiunta diffusione della videoscrittura, ma pur ripubblicandolo numerose

volte fino al 2005, non sottopose mai il suo bestseller a modifiche sostanziali (si limitò invece a prefare nel 1986 l'epigonale seppure onestissimo *Come scrivere una tesi di laurea con il personal computer* di Claudio Pozzoli, BUR); ce lo prova uno sguardo distratto al capitolo VI dell'originale, dove sotto il titolo *La redazione definitiva*, onde meglio simulare la stesura grafica finale della tesi, Eco riproduceva dalla p. 201 in avanti un testo composto non tanto di caratteri "a stampa", ma "a macchina".

Si intuisce così l'altezza spaventosa del tuffo verso quel mondo da non rimpiangere, passata da poco e comunque distantissima; per esempio, chi operava nel campo della filologia classica era solito partire da una raccolta della bibliografia condotta con strumenti già raffinati come l'*Année philologique* – il repertorio basico fondato nel 1928 dal latinista francese Jules Marouzeau; occorreva prendere in mano l'opera e andare a ritroso, tomo per tomo, annotando su schedine manuali quanto se ne ricavava ad uso proprio; nel caso in cui nei lemmari, ordinati per autori o per temi, si trovasse un libro o un articolo attraente (così giudicabile dai titoli / sottotitoli, al massimo da scarse informazioni aggiunte), bisognava procurarsi l'oggetto; ci si muoveva allora per raggiungere l'edificio di una buona biblioteca cittadina, e fortunato era chi entro tempi ragionevoli poteva mettere gli occhi direttamente sugli schedari là dove volumi e riviste già si trovavano; o in seconda istanza sui preziosi (ma incompleti) cataloghi a stampa delle acquisizioni delle altre biblioteche italiane e straniere. Erano così ottenibili le collocazioni di ciò che si cercava, e la caccia continuava subito o con mezzi propri, o con il ricorso al benemerito ufficio del prestito esterno. Il resto del racconto pare intuitivo, ma anche in caso di happy end si può immaginare quali disagi comportasse agli

amanuensi, nel giro di pochi giorni, il lavoro forzato di riproduzione (riduzione, traduzione anche parziale), in pochi giorni, di pagine provenienti dalla Svezia o dall'Olanda, prima dell'avvento delle fotocopiatrici.

Il PC è un "elettrodomestico" trasversale, che in pochi anni ha cambiato le pratiche di lettura e scrittura di gran parte delle persone alfabetizzate, nell'intero globo; ma oltre alle prevedibili ricadute sulla massa della gente comune che (a scopi di lavoro, studio, passatempo) è solita redigere testi propri con programmi di videoscrittura e scaricarne di altrui da internet, conseguenze notevoli – benché spesso inavvertite o non confessate – si sono riflesse sopra l'attività scientifico-accademica quotidiana degli studiosi, anche nelle aree delle discipline umanistiche. Farò un paio di esempi; il primo riguarda il trattamento dei testi già in precedenza pubblicati a stampa dall'autore, ma col tempo bisognosi di revisioni, correzioni, integrazioni ecc.; nell'esperienza di chi ha mandato in tipografia i primi articoletti verso la metà dei '70 del Novecento, per almeno un quindicennio gli aggiornamenti ritenuti più o meno necessari erano introdotti a matita o a penna su "copie di lavoro" personali: simili a dei *pro memoria* in vista delle eventuali riprese di quanto si era scritto – magari in riedizioni o tirature successive, o con destinazione ad altre sedi.

L'esemplare a stampa appariva tutte le volte chiuso, per così dire definitivo, passibile solo di modifiche posticce che richiedevano ulteriori interventi e operazioni meccaniche – così come ciascuna pagina dattiloscritta (ed eventualmente le copie da essa ottenute con la carta carbone) era sottoposta a faticosi rifacimenti che accorpavano le singole correzioni, annotazioni, aggiunte sui margini o interlinea, prima di riassumere un aspetto ordinato. Il sistema

chiedeva agli addetti una pazienza certosina, cura assidua dei particolari minuti, meticolosità mai sufficiente quando sopravveniva la stanchezza che poteva anche produrre il *saut du même au même* a poche righe dalla fine di una intera pagina. Ecco entrare in gioco i più sofisticati mezzi di cancelleria (gomme, correttori scoloranti) per eliminare o coprire i caratteri ad inchiostro errati ed ammendare senza tracce gli sbagli di battitura.

Come avviene da due secoli e più, in modi diversi ma con frequenza crescente, l'innovazione tecnologica ha liberato gli uomini – e potremmo dire gli *animalia* tutti – da gravi fatiche, anzi ha reso inutili molti loro sforzi fisici; se da qualche decennio le macchine stanno contribuendo a varie forme di ricerca gratuita, di conoscenza disinteressata accessibile puramente alle élites intellettuali, non minore è la soddisfazione dei bisogni elementari, con enormi vantaggi per il benessere e la felicità universale. Se suonano false in questi ambiti le litanie contro i danni della 'modernità', più insopportabili appaiono gli epicedi per le corvè, penose e non necessarie, cui si era costretti "quando i mulini erano bianchi"; ho ritrovato in un archivio di vecchi documenti elettronici (scusandomi se le riprendo di peso) queste frasi, stampate una decina di anni fa in una rivista che si chiama «Latina Didaxis» (LATINA DIDAXIS: 2002):

Ora chi facesse un'evocazione nostalgica di carte ingiallite e inchiostri sbiaditi, buchi di tarli e odori di muffa, pagine omesse e schede volanti, si lascerebbe andare solo ad un querulo ed egoistico rimpianto esistenziale, mentre nel pieno della fase di trapasso che viviamo (nelle aree ricche e sagge del mondo, quotidianamente, silenziosamente, i patrimoni di intere biblioteche transitano su nuovi supporti assieme ai loro cataloghi), le risorse dei calcolatori devono piuttosto rasserenare che spaventare quanti hanno a cuore l'obiettivo

autentico: garantire alle future generazioni la nostra eredità di studi, in tempi di ferro quando le opere del passato – come avvenne già sul finire del mondo antico – corrono seri rischi di un naufragio.

Mi rendo conto che un pubblico fatto di colleghi e allievi, di amici e semplici partecipanti, riuniti nel convegno in memoria di Giuseppe Gigliozzi, nonché gli eventuali lettori successivi degli atti, han poco bisogno di sentir lodare i pregi dell'informatica umanistica sul piano della comunicazione scientifica; più utile per tutti sarà allora indicare i rischi da cui occorre guardarsi. Alcune giuste cautele d'altronde mostrava già il maestro della "scuola romana", in un'intervista che sopravvive in rete quale documento di lungimiranza generosa verso i giovani²:

Mi pare assolutamente evidente che questa, come poche altre frontiere, sembrano poter offrire sbocchi professionali agli studenti [...] perché la formazione culturale di un "letterato" unita a una buona conoscenza dei problemi teorici e metodologici [...] costruiscono uno dei miscugli più ricercati. La ricetta è: competenza tecnica unita alla capacità di collocare il fenomeno da trattare nel proprio contesto, di analizzarlo e di trasformarlo in un efficiente oggetto informatico. Come dire: la tecnica è alla portata di tutti, l'analisi raffinata un po' meno. [...] L'utilità delle biblioteche virtuali, che ormai sono molte in tutto il mondo, è incontestabile: si frantumano i tradizionali rigidi spaccati tra campi di studio così diversi, con effetti sulla spazialità, su noi stessi, sulla ricerca. Ci sono però difficoltà nuove che si devono capire. Vanno ricercate su due ordini diversi di motivi. Il nostro progetto si confronta con due tipi di

² Cfr. <http://www.edscuola.it/archivio/stranieri/infoum.html>.

difficoltà: se io digitalizzo un testo, lo faccio ovviamente per poter compiere sul testo elettronico una serie di operazioni che sarebbe impossibile o antieconomico fare utilizzando il testo originale. Dobbiamo garantire la 'conformità' del testo elettronico all'originale. Pensiamo solo alla complessità delle informazioni contenute in un testo letterario (semantiche, filologiche, grammaticali, sintattiche, ecc.) e capiremo quanto complesso sarà il nostro lavoro di descrizione. Non meno importante il problema della portabilità dei documenti, che vuol dire indipendenza dalle piattaforme hardware e software. Non dobbiamo, infatti, rischiare che i nostri dati muoiano assieme a una determinata macchina o assieme a un certo programma, ma dovremo essere sempre in grado di conservare l'informazione, qualunque sia il supporto tecnico. Tutte e due queste opzioni si soddisfano per mezzo di un linguaggio di marcatura che dichiara esplicitamente e in maniera univoca l'informazione che intendiamo conservare, assieme al modo con il quale conserviamo la nostra informazione.

Tutti sanno che la lettura e scrittura a video comportano dei pericoli, legati alla accelerazione di queste pratiche stesse; operazioni in tempi lontani prodotte da movenze lente e meditate (scrivere a mano, ma anche con strumenti meccanici, richiedeva all'individuo un impegno fisico, lo spingeva a dipendere da mezzi e oggetti materiali diversi, lo invitava a gesti ripetitivi per non dire rituali), si sono semplificate e infine deprezzate; questo calo di valore causa inevitabilmente un minor senso di accortezza, fa smarrire la responsabilità ed anzi la solennità di certi atti: quanto più essi sono veloci e volatili – con l'sms si arriva al grado ulteriore rispetto alla e-mail – tanto più le varianti delle scritture e dei pensieri tendono a coincidere: capita che lo scambio di indirizzo, l'errore di inoltrare ai destinatari, ma soprattutto l'avvicinamento nella formulazione

delle idee, ci procurino vari infortuni nella fretta dell'emissione di un messaggio; "non si trattien lo strale / quando dall'arco uscì". E l'istante in cui si preme il tasto di invio accentra in sé la necessaria, obsoleta trafila di lungaggini che comprendeva una rilettura più attenta del foglio.

Si tratta di esperienze comuni, tali da provocare reazioni istintive e rimedi sicuri: non merita soffermarsi su questi aspetti. Ben più serio è il pericolo che nell'attuale fase di trapasso dalla carta stampata ai supporti immateriali – al pari di quanto accadde all'avvento dell'era gutenberghiana – si costituisca una specie di neovulgata dei testi letterari: un guaio tale da richiedere in futuro dei faticosi, e non sempre efficaci, correttivi. Per chi fa il nostro mestiere, l'opportunità di trovare pronti nella rete, o su repertori in CD-ROM, intere biblioteche plurilingui donde estrarre i brani di testo su cui operiamo, e che poi eventualmente adattiamo inserendole nei nostri saggi, appare impagabile; ma c'è qualche insidia nascosta.

Ogni giorno, nella consuetudine dell'insegnamento e della ricerca, un antichista si serve di banche-dati specialistiche consultabili online; ce ne sono varie, provengono in larga parte da iniziative di matrice accademica nordamericana e contengono testi di epoca e genere diversi, più o meno rari e conosciuti, più o meno accurati nella trascrizione e nel trattamento grafico; in particolare, qualunque utente di internet è da anni messo in grado di leggere gratuitamente, e scaricare se lo desidera, una delle fonti più preziose per la nostra conoscenza della cultura letteraria romana: i *Saturnalium libri*, che furono composti da Macrobio nel quarto decennio del V secolo; basta indirizzarsi al portale *Lacus Curtius*

curato da Bill Thayer (University of Chicago, ne è disponibile perfino la versione in italiano)³.

Nel preparare queste pagine, mi occupavo contemporaneamente di una questione di prosopografia relativa ad un erudito scrittore latino di età imperiale. Sereno Sammonico (questo il suo nome) trattava di prelibatezze culinarie e delle spese che i ricconi di età repubblicana affrontavano per procurarsi l'*acipenser* – cioè lo storione. Al tempo di Plinio e di Traiano, il prezzo di quel pesce al mercato era diminuito, ma le cose erano destinate a girare nuovamente più tardi. Ecco la forma in cui si legge in rete il passo interessato (Macr. Sat. 3, 16):

5 Nec infitias eo temporibus Traiani hunc pisces in magno pretio non fuisse, teste Plinio Secundo qui in Naturali historia, cum de hoc pisce loqueretur, sic ait: Nullo nunc in honore est, quod equidem miror, cum sit perrarus inventu. 6 Sed non diu stetit haec parsimonia. Nam temporibus Severi principis, qui ostendebat duritiam morum, Sammonicus Serenus, vir seculo suo doctus, cum ad principem suum scriberet faceretque de hoc pisce sermonem, verba Plinii quae superius posui praemisit et ita subiecit eqs.⁴

In queste sue poche righe, il testo (che forse per motivi di copyright riproduce una veneranda stampa tedesca di metà

³ http://penelope.uchicago.edu/Thayer/L/Roman/Texts/Macrobius/Saturnalia/3*.html.

⁴ "Non intendo negare che ai tempi di Traiano questo pesce fosse meno caro, come attesta Plinio Secondo che parlandone nella *Storia naturale* dice: 'Ora non è affatto apprezzato, e me ne stupisco, perché è raro a trovarsi'. Ma questa parsimonia non durò a lungo; ai tempi dell'imperatore Severo, che ostentava rigidità di costumi, Sammonico Sereno, un dotto di quell'età, scrivendo all'imperatore venne a parlare di questo pesce e, dopo aver citato le parole di Plinio riferite sopra, aggiunse ecc."

Ottocento) nasconde alcune divergenze rispetto alle edizioni critiche pubblicate negli scorsi decenni: per esempio, alla ultima e recentissima (KASTER: 2011) di Robert A. Kaster, che scrive così:

5 nec infitias eo temporibus Traiani hunc piscem in magno pretio non fuisse, teste Plinio Secundo, qui in Naturali historia cum de hoc pisce loqueretur, sic ait: "nullo nunc in honore est, quod equidem miror, cum sit rarus inventu." 6 sed non diu stetit haec parsimonia. nam temporibus Severi principis, qui ostentabat duritiam morum, Sammonicus Serenus, vir saeculo suo doctus, cum ad principem suum scriberet faceretque de hoc pisce sermonem, verba Plinii quae superius posui praemisit et ita ipse subiecit eqs.

In un paio di casi si tratta di varianti adiafore. All'interno della citazione letterale pliniana, mediata da Sammonico e riportata da Macrobio, *cum sit rarus inventu* è stato ripristinato da Nino Marinone (MARINONE: 1967) sulla base del confronto di Plin. nat. 9, 60, laddove la paradosi dei codici dei *Saturnalia* è chiaramente guasta (*parvus* ω) e Ludwig von Jan (1852) aveva appunto congetturato *perrarus*. Ancora, in *praemisit et ita subiecit*, proprio alla fine della pericope riportata, si è ripristinato il pronome *ipse* – un elemento logicamente e grammaticalmente non necessario al periodo, omissa da tutti gli editori prescientifici.

Più insidioso per i lettori è un terzo caso di difformità, certamente introdotto al momento della riscrittura in versione elettronica. La scorrettezza si cela nella frase [...] *Severi principis, qui ostendabat duritiam morum*, dove il verbo – per errore non tanto di un programma di riconoscimento automatico dei caratteri, ma del moderno copista 'amanuense' – si genera dall'incrocio tra due forme di imperfetto: *ostendebat* (da *ostendere*) e *ostentabat* (da *ostentare*,

derivato intensivo del precedente), offerta dalla universalità dei codici e dalle stampe.

Si capisce come il refuso che interessa l'unica (oggi e forse per lungo tempo) redazione online disponibile, quindi potenziale sorgente di vulgata, rischia di estendersi fuori controllo in ogni futura ripresa del testo latino; a meno che il guasto non venga segnalato al curatore del sito – come peraltro sin dalla pagina introduttiva egli ci raccomanda⁵ con la calda esortazione "Feedback and Corrections Gratefully Accepted!" – e rettificato di conseguenza.

Su questi aspetti la vigilanza deve essere mantenuta alta, non c'è dubbio; ma per comprendere appieno i benefici di cui godiamo oggi rispetto alle regole tecniche della stampa è utile guardare ad un altro esempio offerto dallo stesso testo tardoantico; sia permesso ricavarlo da un saggio comparso nel lontano 1997 sulle pagine di una rivista accademica nazionale, ben diffusa tra gli specialisti; titolo e sottotitolo dicono abbastanza dell'argomento trattato: *Seneca e il copista infedele. Il testo delle Ad Lucilium nelle rielaborazioni di Macrobio*, "Paideia" 52, pp. 191-223. Il compilatore dei *Saturnalia* – come usava all'epoca – spesso riporta parola per parola larghe sezioni di opere precedenti, anche senza citare la fonte: una forma di appropriazione che prepara la perdita di autorialità che caratterizza molta produzione letteraria medievale. Nei capitoli iniziali del dialogo, dove tema è l'origine del ciclo festivo di dicembre, sono rielaborati materiali tratti da Seneca; in particolare, si reimpiega la celebre epistola 47 in cui il filosofo esaltava il principio

⁵ <http://penelope.uchicago.edu/Thayer/E/HELP/corrections.html>.

della eguale dignità per tutti attraverso atteggiamenti di umanità dei padroni nei confronti degli schiavi (*Serui sunt; immo homines ...*).

Altre volte Macrobio si diverte a imitare con parole sue lo stile del pensatore stoico, esprimendo opinioni che a lui comunque potevano riferirsi; ed è questo il caso che segue. Da secoli, in qualsiasi edizione a stampa dei *Saturnalia*, all'altezza di 1, 11, 9 troviamo scritto:

[...] et certe nulla seruitus turpior quam uoluntaria. At nos iugo a Fortuna imposito subiacentem tamquam miserum uilemque calcamus, quod uero nos nostris ceruicibus inserimus non patimur reprehendi.

La lezione *inserimus* è accolta senza problemi da tutti gli editori, ma il senso della frase sorprende e nei fatti spiazza chiunque debba volgere la frase in un'altra lingua⁶; noi il peso gravoso, l'onere molesto, ce lo "imponiamo" e non "inseriamo" sul collo: bisogna emendare il verbo in *inferimus*, tanto più che la confusione tra le due forme nelle tradizioni manoscritte è ben documentata dal lemma del

⁶ Riporto alcune fra le più diffuse traduzioni oggi circolanti; N. Marinone (Torino, 1967): "[...] e certo nessuna schiavitù è più turpe di quella volontaria. Gli è che noi calpestiamo come miserabile e vile chi soggiace al giogo impostogli dalla sorte, e non sopportiamo di essere rimproverati per quello che *ci siamo posti noi stessi sulla nuca*"; J. F. Mesa Sanz (Madrid, 2009): "Con seguridad, mo hay esclavitud más vergonzosa que la voluntaria. Ahora bien, nosotros, impuesto el yugo por Fortuna, pisoteamos al caído como si fuera desgraciado y vil; no soportamos, en cambio, ser reprehendidos por lo que argamos sobre nuestras cervices"; R. A. Kaster (Cambridge Ma., London, 2011): "And surely no slavery is worse than the one we chose for ourselves. Yes, we trample on the man who lies fallen beneath the yoke that bad luck has put upon him, as though he were wretched and of no account – but we don't stand criticism for the yoke that *we've placed on our own necks*".

Thesaurus linguae Latinae (VII/1, 1869,50 s. [Stiewe]), mentre l'identica espressione *ceruicibus inferre* (usata in senso morale) ritorna nel contemporaneo poeta Aviano, fab. 37, 15 (*ibidem*, 1375,38 [Hofmann]). Ammesso che la universalità dei codici macrobiani rechi davvero *inserimus* (e dobbiamo crederlo solo *ex silentio* degli apparati critici), *inferimus* ha poi il sostegno di una autorevolissima tradizione indiretta, rappresentata da Giovanni di Salisbury; come infatti dipendevano da Seneca le pagine dei *Saturnalia*, così dopo sette secoli queste ultime finivano reimpiegate a loro volta nel capitolo 8, 12 del *Policraticus*, sotto il titolo 'Quanta humanitate cum seruis uiuendum sit' (WEBB, p. 307-309; ciò che interessa si trova a p. 308, 3; il vecchio testo del Migne si può controllare in rete⁷).

Quale morale trarre, da queste storie di minuzie (paleo- e tipografiche)? Che la *vulgata* è dura a morire, la sua inerzia pervicace: ma essa non si lega alla innovazione tecnica dell'editoria digitale – che di contro permette il ritocco del testo in qualunque momento successivo alla sua pubblicazione; certo non deve giudicarsi adesso più dannosa che in passato: e proprio l'ultimo esempio mostra che il rischio non riguarda soltanto il presente e il futuro della filologia ecdotica, ma si è trascinato lungo almeno tutti i secoli della stampa, dagli incunaboli in qua. Il tipico scambio *s / f*, banale errore di trascrizione della minuscola carolina (dove i caratteri sono quasi identici, e distinti solo per un trattino orizzontale), è passato indenne attraverso un discreto numero di revisioni critiche del testo, nonché – si presume – di autopsie di manoscritti.

⁷ <http://monumenta.ch/latein/>.

Un'osservazione per concludere, riguardo alle scelte di indirizzo generale che toccano le nostre discipline. Alla luce di tante esperienze degli ultimi decenni, a volte sviluppate sino a livelli eccellenti di funzionalità, a volte lasciate inconcluse nonostante il loro provato valore, sembra chiaro che la via regia per andare avanti è la creazione di grandi *corpora* digitali, o in molti casi il perfezionamento di archivi specializzati già esistenti. Incentivare e finanziare iniziative di questo tipo – che per loro natura instradano alla formazione nel lavoro di équipe – è la 'politica' più giusta e più utile nei confronti dei giovani, in questo momento; forse l'unica davvero remunerativa in assoluto, nei tempi lunghi. Accade di pensarlo quando si guarda alle energie spese dall'editoria cartacea tradizionale per aggiungere ai cataloghi delle collezioni l'ennesima prova di edizione critica sopra testi (antichi e moderni, non solo grandi 'classici') ormai largamente consolidati, persino nei luoghi in cui la lezione permanga dubbia o insoddisfacente. Dopo secoli e secoli di attività interpretativa ed eventualmente emendatoria, vien da chiedersi a quali sviluppi attendere ancora sul piano della *constitutio textus*, quando si escluda l'apporto di testimonianze manoscritte ulteriori. Forse nulla, che non sia l'illusione di attingere la mitica 'volontà dell'autore', o peggio la vanità di identificarsi con lui e riscrivere al suo posto, per congettura. A queste debolezze individuali si contrappone una prospettiva di studi che mirino apertamente alla ricostruzione storico-filologica; che diano valore al testo per le sue interne vicende, seguendone gli itinerari e le fasi di evoluzione nello spazio e nel tempo, ma anche per apprezzarne gli effetti esterni sulle diverse culture ed ambienti di ricezione, cioè sugli autori successivi e la totalità della produzione letteraria.

Domande teoriche sul concetto di edizione (nel nome di Giuseppe Gigliozzi)

Raul Mordenti

Considero questa un'occasione rara e preziosa, per discutere liberamente, senza troppi vincoli accademici ma con colleghi e amici e maestri estremamente qualificati; interpreto questa possibilità come un ennesimo dono che ci viene da Giuseppe, dato che è il ricordo di lui che riunisce qui – dieci anni dopo la sua scomparsa – tante persone, e così diverse. Vorrei dunque utilizzare questa occasione per porre alcune questioni teoriche, che a me sembrano di fondo, anche se il mio personale livello di riflessione non mi consente ancora di dare a questi interrogativi risposte esaurienti e soddisfacenti. Ma come diceva Arnaldo Momigliano, ci sono fasi in cui per un ricercatore la cosa più importante è porsi le domande giuste, e questo è ancora più importante che non cercare di darsi comunque le risposte, magari sbagliate.

1. Le domande su cui mi interrogo, e vorrei interrogarvi, ruotano attorno al concetto delle "due fasi"; è un concetto che mi è particolarmente caro, che è ben noto ai miei amici, oltre che ai miei

studenti, e di cui mi è capitato altrove e più volte di parlare e scrivere; mi scuso dunque della ripetitività.

A me pare che le cosiddette "due fasi" caratterizzano la storia delle invenzioni umane: in una "prima fase" la nuova invenzione viene chiamata a rispondere a domande vecchie, cioè a domande sorte nell'ambito del quadro epistemologico dato; solo in una "seconda fase" si comprende che la nuova invenzione è veramente *una cosa nuova*, cioè che essa non si limita a risolvere i problemi del vecchio assetto euristico ma che determina un campo di problemi assolutamente nuovo, spesso del tutto impreveduto, e che dunque la sua importanza non consiste tanto nel risolvere i vecchi problemi, quanto – per così dire – nella capacità di renderli semplicemente insensati, di spazzarli via, determinando (e però anche risolvendo o promettendo di risolvere) dei problemi del tutto nuovi.

Gli esempi che allego sono noti, sono numerosi, ma si potrebbero facilmente moltiplicare (anzi questa è una prima domanda che pongo: vedere assieme se vi sono altri casi in cui questo schema si possa applicare, e se esso ha veramente un significato e una validità generale).

Si potrebbe – per esempio – parlare in questi termini della cosiddetta "scoperta dell'America": in tal caso la "prima fase" sarebbe la scoperta di una via occidentale alle Indie, mentre la "seconda fase" sarebbe rappresentata dalla possibilità di entrare in contatto (e, essendo europei, in possesso) di un nuovo continente del tutto impreveduto. Così come – lo sappiamo bene –, la stampa è stata davvero troppo a lungo (per cinquant'anni almeno) solo una pedissequa imitazione del manoscritto. Appartiene ancora alla "prima fase" il fatto che si sia continuato al principio dell'era

gutenberghiana ad usare le abbreviazioni, ormai prive di senso, dato che nella nuova procedura produttiva il peso del tempo-costo del copista, che le motivava, era adesso diventato un fattore relativamente trascurabile a fronte della facile moltiplicazione meccanica delle copie, mentre erano diventati decisivi altri costi (ad esempio, quello della carta); per restare ancora alle abbreviazioni: la stampa le rendeva addirittura dannose, a causa della scomoda e costosa necessità di fondere e utilizzare un numero maggiore di glifi, aggiungendo a quelli alfabetici anche quelli abbreviativi, mentre la logica interna dell'invenzione di Gutenberg era, evidentemente, ridurre al minimo il numero dei caratteri di piombo da utilizzare (e non per caso il *set* delle lettere dell'alfabeto in uso nell'era chirografica con la stampa si ridusse, e si normalizzò). Ma ci volle del tempo per accorgersi di ciò che a noi adesso pare del tutto evidente. Analogo discorso si potrebbe fare per altri aspetti del libro a stampa, che solo a fatica (e non subito) si differenziarono, liberandosene, dal vecchio modello del manoscritto, come – ad esempio – la *mise en page* o il formato dei libri che comportava spreco di carta, o la presenza anti-economica di larghi margini bianchi, o la numerazione irrazionale dei quaderni invece che delle pagine, e così via. La fotografia, è stata considerata a lungo, assai riduttivamente, solo un quadro senza pennelli, che conservava dunque del quadro – per dir così – la sintassi, cioè la locazione, la posizione del soggetto, l'inquadratura, ecc.; solo tempo dopo qualche genio (penso a Nadar) si accorse che si poteva fotografare all'aperto, che si poteva fotografare in un istante anche soggetti in relativo movimento, che si poteva fotografare perfino là dove l'occhio umano non vedeva (ad esempio nei sotterranei della costruenda metro di Parigi, ecc.). La

radio è stata considerata all'inizio come un sistema di comunicazione senza fili "da uno a uno", senza capire subito che il suo vero e straordinario potere consisteva invece nel poter trasmettere da uno a molti, e anche il telefono fu usato a lungo del tutto irrazionalmente (c'è stato perfino il "teatrofono" usato da Proust, che consentiva di sentire Sarah Bernhardt standosene a casa, appesi alla cornetta del telefono). In qualche mio libro ho pubblicato un'immagine comica, o tragicomica, ma particolarmente significativa: una delle prime automobili che nascondeva il motore dentro un finto cavallo, per sembrare ancora una carrozza. E, ripeto, l'elenco potrebbe facilmente (e forse inutilmente) allungarsi.

Ebbene mi sembra che nel campo dell'informatica umanistica si tratta di passare, e forse stiamo passando – senza accorgercene abbastanza e dunque senza affrettare il passaggio come sarebbe possibile e necessario – dalla "prima fase" alla "seconda fase". Di certo anche della nostra "prima fase" qualche postero ingeneroso sorriderà, esattamente come noi sorridiamo delle stampe che imitavano i manoscritti o delle automobili con la testa impagliata di un cavallo.

2. Mi domando: un *surplus* di riflessione teorica sui fondamenti, del tipo per intenderci di quella a cui Tito Orlandi ci chiama instancabilmente, può aiutarci ad essere più coscienti di questo passaggio di fase, e ad accelerarlo? Domandiamoci infatti: che cosa è che *non funziona* nella "prima fase", al punto da renderla agli ingenerosi posteri addirittura ridicola? Io direi che il limite della "prima fase" (e di tutte le "prime fasi") sia un limite di *conservatività teorica* (insisto: teorica, non tecnologica!), cioè si ritarda a considerare la novità per ciò che effettivamente essa è teoricamente, e così, e per

questo, si tende a versare – per così dire – il vino nuovo negli otri vecchi. Penso ad esempio – abbiamo qui Gino Roncaglia che è maestro dell'argomento – che questo sia stato evidente nella vicenda degli ebooks propriamente detti o di prima generazione, in cui si imitava ancora il libro in quanto oggetto, considerando cioè ancora il nuovo oggetto elettronico *un contenitore di testo*, solo reso digitale da analogico e cartaceo che era (dunque con molti sprechi e contraddizioni), mentre è evidente che la "seconda fase" è qui rappresentata dalla possibilità di accedere più o meno direttamente, tramite la rete, più o meno a tutti i libri del mondo, cioè a tutti i libri che stanno in rete; ma in questa "seconda fase", come si vede, l'ebook non "contiene" un bel nulla, esso è diventato solo *un accesso* portatile ad internet e un'interfaccia facilitante, cioè è diventato davvero tutta un'altra cosa rispetto al libro, o meglio rispetto alla vecchia *idea* di libro. Se questo è dunque il quadro generale di riferimento delle mie domande, passo ora a formularne alcune (essenzialmente due, anche se suddivisibili in molte altre sotto-domande): la prima riguarda le modalità di produzione e fruizione del libro (e mette dunque in questione l'editoria); la seconda riguarda il concetto di edizione critica.

3. La produzione, la gestione e la circolazione dei testi informatici è – mi sembra – è ancora affidata ad una configurazione tipica del testo a stampa (diciamo: gutenberghiana) che, probabilmente, non ha più alcun senso nel tempo della rete. In particolare cambia del tutto (e forse non lo abbiamo capito abbastanza!) il ruolo svolto dall'editore (inteso qui per ora come *publisher*, non come *editor*): centrale nella configurazione gutenberghiana, del tutto marginale, se non addirittura superfluo e

insensato, nella configurazione segnata dall'informatica. Un vecchio adagio (che ho imparato da un'amica editrice, Luisa Capelli) diceva che "L'editore è colui che non scrive i libri, non li stampa, non li distribuisce, non li vende, ma... fa tutto il resto". Cosa faceva in realtà l'editore, l'editore gutenberghiano? Egli *mediava* fra il testo, il suo autore, e il pubblico, e per questa attività di mediazione, davvero cruciale, egli riceveva una giusta remunerazione (la quale anzi svolgeva un ruolo di innesco e motivazione dell'intero processo di produzione-distribuzione-fruizione del testo, cioè era a sua volta – la remunerazione – centrale ed ineliminabile). Queste attività di mediazione erano molte, ne elenco qui (sommariamente) solamente sei: 1) la *selezione* del testo da pubblicare, o addirittura la sua progettazione nelle collane (in questi casi l'editore committente era in realtà un co-emittente del testo, qualcosa di simile, e forse qualcosa di più, di un autore); 2) la connessa *garanzia* culturale e pubblica della qualità, perché non era la stessa cosa pubblicare da un editore prestigioso o da uno, come si diceva una volta, "a pagamento" (ora non si dice più perché sono quasi tutti a pagamento); 3) l'attività di revisione redazionale del testo, più o meno ingente ed accurata a seconda della qualità dell'editore (e ciò rendeva spesso l'editore, o i suoi redattori, dei veri e propri co-autori: si pensi al ruolo svolto per Einaudi da redattori come Italo Calvino o Elio Vittorini); 4) il sostegno e la promozione del libro presso il pubblico dei lettori, con campagne culturali fatte essenzialmente di recensioni e di presentazioni, ecc. (che un tempo l'editore stesso organizzava); 5) la diffusione del libro e la garanzia della presenza del libro nelle librerie, a cui l'editore provvedeva o direttamente o tramite agenzie di distribuzione (un discorso questo particolarmente critico che ci porterebbe lontano); 6) la

conservazione del libro nel proprio catalogo, ciò che rendeva un libro accessibile nel tempo ai lettori (questo era peraltro un indice di qualità editoriale perché testimoniava che quei libri erano destinati a durare, rendeva l'editore – appunto – un "editore di catalogo").

Se ci riflettiamo, *nessuna* di queste funzioni viene oggi più svolta dagli editori, almeno dalla maggior parte degli editori (le eccezioni sono sempre meno): 1) non la selezione e progettazione del volume (che è anzi di solito è proposto o insistentemente offerto dall'autore stesso); 2) non la garanzia culturale di qualità, perché come – si è detto – sempre più spesso c'è un pagamento da parte dell'autore, e se si riceve denaro dall'autore non si va poi troppo per il sottile sul testo proposto per la pubblicazione; si noti che fra l'altro tale pagamento anticipato esime del tutto l'editore dalla necessità di rientrare dei costi tramite la vendita del libro, e gli risparmia in particolare la costosa distribuzione (e gli editori che, come la meritoria Meltemi, rifiutano per principio qualsiasi pagamento o "contributo" da parte degli autori hanno incontrato, non per caso, gravi difficoltà economiche); 3) non l'attività di revisione redazionale, che anzi le redazioni sono la prima cosa che è stata – come si dice – "esternalizzata", cioè distrutta (con un danno incalcolabile di esperienza e saperi editoriali, che le redazioni si formavano nel corso degli anni); 4) non le recensioni o le presentazioni, che l'autore si deve procurare da solo (se e quando ci riesce...); 5) non la distribuzione e la presenza nelle librerie, perché i piccoli editori ci rinunciano ormai del tutto, a causa dei costi esorbitanti (e delle posizioni dominati in libreria dei grandi editori), e non a caso sempre più si affidano alla rete internet per cercare di vendere direttamente ai lettori; 6) non infine la conservazione nel tempo del titolo e della sua accessibilità sul mercato, perché il

catalogo e il magazzino rappresentano un costo, e non è affatto infrequente che l'autore si veda offrire dall'editore il suo stesso testo, a prezzo scontato (ma è un prezzo pagato per la seconda o terza volta!), avendo come alternativa la minaccia del macero, psicologicamente intollerabile e angosciante per ogni autore (sulla base della sua esperienza di impiegato nell'editoria, Umberto Eco ha descritto da par suo questo modo di procedere nel *Pendolo di Foucault*).

4. Non sarebbe elegante fare dei nomi, e non è questa la sede; ma sembra evidente che se questa configurazione di mediazione fra autore, testo e pubblico dei lettori, che l'editore svolgeva, entra in crisi nel suo complesso (e l'informatica è parte essenziale di questa crisi, o collasso che dir si voglia), allora non si vede poi perché debba essere conservato proprio e solo un punto di essa, cioè la remunerazione dell'editore per una funzione di mediazione che egli di fatto non svolge più. Detta ancora più brutalmente: perché mai io dovrei remunerare un editore che non sceglie il mio libro né lo progetta, che non lo convalida culturalmente col suo prestigio, che non ne opera alcuna revisione editoriale degna, che non garantisce recensioni e presentazioni né la presenza del mio libro nelle librerie, e neppure la sua durevole reperibilità nel proprio catalogo e magazzino? Senza dire che a volte anche la composizione tipografica del libro, dai titoli correnti all'indice all'impaginazione delle note, è affidata al lavoro (gratuito) dell'autore stesso (a me personalmente è successo, e credo di non essere il solo). In questi ultimi casi il problema non è il 10% che va, o dovrebbe andare, all'autore, il problema è capire perché mai il 90% debba andare all'editore!

Nella vecchia configurazione, che abbiamo definito gutenberghiana, l'autore partecipava, tramite l'editore, alla remunerazione con il cosiddetto copyright (che forse sarebbe ormai tempo di chiamare più precisamente "diritto d'editore" e non più "diritto d'autore"!)). Dunque il discorso evidentemente riguarda e coinvolge anche il diritto d'autore, e a proposito del copyright io mi limiterò a dire alcune cose che mi sembrano ovvie, e che non vorrei sembrassero provocatorie (ma viviamo in un tempo infelice in cui le ovvietà suonano spesso come provocazioni):

1) anzitutto che il copyright non è sempre esistito, né sempre esisterà; Petrarca – per esempio – non si è mai sognato di campare coi diritti delle sue opere, ma esistevano al suo tempo forme di remunerazione pubblica e sociale dell'ingegno (e di garanzie dell'autorialità morale) che ormai conosciamo abbastanza bene, e che forse hanno oggi qualcosa da insegnarci; avanzo l'ipotesi che sia necessario tornare in modi storicamente adeguati a queste forme pre-capitalistiche di sostegno pubblico e diretto ai produttori di libri (e non a chi li edita), forse non attraverso il conferimento di rendite ecclesiastiche e di ordini minori (di cui usufruirono sia Petrarca che Boccaccio) però attraverso un adeguato finanziamento pubblico di queste attività autoriali, almeno di quelle frutto di ricerca svolta dalle università. Forse per la letteratura cosiddetta "di intrattenimento" (strana e labile definizione!) il discorso è diverso, ma faccio notare che quasi *nessuno* scrittore del Novecento vive dei diritti d'autore: Montale, che era Montale, viveva del "Corriere della Sera", Moravia che era Moravia delle recensioni cinematografiche sull'"Espresso", Vittorini e Calvino non poterono mai rinunciare al loro lavoro di redattori editoriali né Primo Levi al suo lavoro di chimico, e così via; per non dire delle difficoltà economiche gravissime che accompa-

gnarono fino alla morte (per fare un solo esempio) Elsa Morante, non solo la scrittrice italiana più grande del secolo scorso, ma anche fra le più vendute;

2) che il *copyright* è strettamente legato alla forma-merce del testo, cioè all'oggetto-libro a stampa ed all'industria editoriale (dunque, se ci pensiamo, è durato pochi secoli), e che farsi pagare un non-libro informatico incontrerà problemi analoghi a quelli già incontrati dall'industria discografica nel cercare di farsi pagare dei non-dischi, cioè i file di musica. I migliori auguri a quelli che ci hanno provato, ci provano e ancora ci proveranno (io credo invano), a condizione che per trarre profitto dalla rete costoro non riescano a mettere qualche mordacchia al web;

3) che, infine, l'essenziale è per noi garantire il rispetto dell'autorialità – diciamo così – "morale" (e questo sembra possa essere largamente garantito dalle modalità cosiddette *copyleft* e dal sistema *Creative Commons*), ma tutt'altra cosa è la bizzarra idea che da ogni atto di lettura del suo testo un autore debba trarre un profitto;

4) che, in ogni caso, sembra *assolutamente intollerabile* il fatto che le università non abbiano i soldi per poter leggere e far leggere ai propri studenti nelle loro biblioteche i saggi o gli articoli o le ricerche prodotti dai loro stessi professori; questi lavori sono ceduti dagli autori gratuitamente alle riviste scientifiche, e dalle stesse riviste sono fatti pagare a costi salatissimi, anzi tanto più salati quanto più la rivista è importante e dunque la lettura dei suoi saggi indispensabile. Qui occorre dunque cambiare, e radicalmente. Questa follia è tanto più insopportabile se la collettività ha già pagato quel prodotto dell'ingegno (chiamiamo immodestamente così le nostre opere), anzi le ha pagate due volte, sia perché l'autore

universitario già gode di uno stipendio in quanto dipendente dell'università pubblica, sia perché spessissimo (almeno in Italia) tali pubblicazioni sono pagate agli editori coi fondi pubblici della ricerca, i quali dunque spesso finiscono per pagare editori privati (e i libri così prodotti sono destinati a finire ammuccati nei corridoi dei nostri Dipartimenti). D'altra parte, credo che non ci siano uomini o donne di cultura degni di questo nome che – chiamati a scegliere – non preferiscano mille volte che la loro opera sia letta da migliaia di persone in tutto il mondo, piuttosto che trarne qualche (di solito, nei nostri casi, miserabile) profitto. Se l'università paga il mio lavoro di ricerca, perché mai il prodotto di quel lavoro, per ipotesi un libro, deve arricchire il mio editore, e magari anche me, e non invece l'università che mi stipendia? Insomma la gratuità e il carattere pubblico dell'università a me sembrano strettamente legate: *simul stabunt e simul cadent*, e proprio perché siamo in una fase in cui è attaccato direttamente e violentemente (e, ahimè, non senza visibili risultati) il carattere pubblico dell'università, io credo che non possiamo più permetterci il lusso di trascurare questo punto vitale. Peraltro oso dire che a me sembra che anche la rete sia essenzialmente gratuita *per sua natura*, e che gli sforzi di far pagare la rete somiglino, appunto, agli sforzi di coloro che imitavano supinamente con la stampa i manoscritti, o che mettevano teste di cavallo impagliate sulle automobili per farle somigliare alle carrozze. Dunque l'incontro fra università pubblica, informatica e *open access* mi sembra che sia nella natura delle cose, e nella natura delle cose mi sembra sia che tale incontro avvenga sotto il segno della gratuità.

Insomma, credo che l'informatica ci consenta e ci costringa (al tempo stesso le due cose! ci consente e dunque ci costringe) a guardare alla cultura per quello che effettivamente essa è, cioè come

a un bene comune (e dico comune non privato; comune non statale!), cioè un bene le cui caratteristiche essenziali sono la *diffusione* e la *condivisione*, e a cui non è possibile applicare il principio di esclusività ("o questo bene è mio, *oppure* è tuo, se è mio non è tuo": un principio rozzo tratto, a ben vedere, dal modello arcaico costituito dalla proprietà della terra). Esattamente come succede per un tramonto o per un paesaggio o per l'acqua e l'aria pulita, il fatto che altri ne goda non impedisce affatto che ne goda io, e lo stesso è per la cultura; si potrebbe anzi sostenere la tesi inversa, cioè che della cultura si gode tanto più quanto più essa è diffusa e condivisa, e che tale generale diffusione e condivisione rappresenta un interesse primario, e non comprimibile, di una società che si voglia dire civile.

5. Vengo allora alla seconda domanda – che è quella che mi interessa di più – che riguarda il concetto stesso di edizione critica. Formulerei la domanda in questi termini: nel concetto vigente di edizione quanto è davvero essenziale al concetto stesso, in sé considerato, e quanto invece è debitore della stampa, intendo dire dell'epistemologia di Gutenberg non solo della sua tecnologia? Anche se fra le due cose, l'epistemologia di Gutenberg e la sua tecnologia, c'è evidentemente un rapporto, a noi interessa e deve interessare più la prima, l'epistemologia (la quale, come abbiamo visto all'inizio di questo nostro ragionamento, ancora rischia di condizionarci e frenarci trattenendoci nella "prima fase") che non la seconda, la tecnologia, la quale muore da sola, come tutte le cose che meritano di morire. Penso (schematizzando al massimo, ancora una volta, per le caratteristiche di questo mio sommario contributo) soprattutto a cinque caratteristiche dell'edizione critica che conosciamo, e che – per dir così – abbiamo ancora nelle nostre teste a

comandarci (magari senza che noi stessi ce ne accorgiamo). Mi domando, e vi domando, se queste cinque caratteristiche possano ancora essere vigenti di fronte di un'edizione informatica dei testi.

1) Penso in primo luogo al concetto di *definitività* dell'edizione (*ne varietur!*), e alla nettissima distinzione fra lavori preparatori (che restano, *et pour cause*, inediti nella configurazione gutenberghiana) e l'edizione vera e propria, cioè definitiva, del testo.

2) Penso, in secondo luogo, all'idea (legata al primo punto, ma logicamente da distinguere) che il testo da editare sia, e sia sempre, e debba essere sempre, uno e uno solo.

3) Penso, in terzo luogo, alla netta gerarchia fra testo e apparato, cioè fra la soluzione prescelta dall'editore in base allo stato (per definizione sempre provvisorio) della ricerca filologica, e le lezioni rifiutate, confinate nelle note.

4) Ancora, mi domando quanto sia sostenibile, in ambiente informatico, la nettissima gerarchia di importanza scientifica (imposta dall'edizione gutenberghiana) fra l'edizione critica propriamente detta e l'edizione diplomatica, o diplomatico-interpretativa (o, come noi ormai diciamo, EDIC, Edizione Diplomatico-Interpretativa Codificata).

5) Infine domando se sia ancora possibile pensare alla esclusiva "autorialità" dell'editore (chiamiamola così), cioè al carattere assolutamente individuale, e non collettivo del lavoro di edizione (superfluo dire che qui, e d'ora in poi, si intende editore come *editor* e non più come *publisher*).

Ora, io credo che si possa avanzare almeno l'ipotesi (dunque la domanda che mi e vi rivolgo) che una modalità propriamente informatica di edizione (e non ripeto qui la distinzione proposta altrove fra la modalità "propriamente informatica" e quella

"meramente digitale" dell'edizione) metta in crisi, uno per uno, *tutti* questi cinque punti appena richiamati. Esaminiamoli partitamente.

1) Al concetto di definitività dell'edizione (*ne varietur*), e alla nettissima distinzione fra edizione critica da pubblicare e lavori preparatori che restano inediti nella configurazione gutenberghiana, si oppone, mi sembra, la modalità informatica dell'edizione; qui ad una possibilità tecnologica (correggere e modificare l'edizione senza dover sopportare quasi nessun costo, ben diversamente da quanto avveniva nella configurazione gutenberghiana), si aggiunge utilmente una opportunità scientifica, che definirei quasi di deontologia della ricerca consentita dall'informatica; mi riferisco alla possibilità – e dunque, in un certo senso, alla necessità – di rendere disponibile il lavoro in ogni momento del suo farsi, affinché la comunità scientifica possa intervenire e discutere e, ove necessario, correggere. Faccio notare che questa modalità è già oggi largamente praticata dai nostri colleghi scienziati, e un settore avanzatissimo e fortemente concorrenziale come quello della fisica delle particelle mette in rete (nella forma di "archivi" online accessibili a tutti e sempre più aggiornati da tutti e da ciascuno) spunti, ipotesi di ricerca, pre-print e quant'altro; è stata questa, ad esempio, la modalità di comunicazione della recente possibile straordinaria scoperta del neutrino più veloce della luce, nota anche al grande pubblico per "il tunnel del ministro Gelmini", che collega direttamente il CERN di Ginevra al laboratorio del Gran Sasso.

2) All'idea (legata al primo punto, ma logicamente da distinguere) che il testo da editare sia, e sia sempre, e debba essere sempre, uno e uno solo, si oppongono obiezioni antiche e profonde della filologia, che ora ridiventano del tutto attuali (forse perché tali

obiezioni diventano ora, e solo ora, la base di possibili pratiche ecdotiche e non solo stimoli teorici affascinanti). Mi basta qui richiamare l'idea di Giorgio Pasquali che non vi sia stato sempre un archetipo e uno solo (archetipi in movimento, tradizioni tutte genuine e però differenziate fin dalla loro origine, per non dire del fenomeno della contaminazione, delle varianti d'autore ecc.), e non c'è dubbio che questa idea sia oggi praticabile anche in sede di edizione, dato che diventa oggi del tutto possibile produrre edizioni che tengano conto concretamente di questa *mouvance* testuale.

3) Anche alla netta e irreversibile gerarchia fra testo e apparato (cioè fra la soluzione prescelta dall'editore, in base allo stato per definizione sempre provvisorio della ricerca filologica, e le lezioni rifiutate confinate nelle note) si oppongono in modo – di nuovo! – convergente sia la più illustre teoria filologica (penso a Contini) che le possibilità offerte dall'informatica; l'edizione informatica può effettivamente fare sì che le lezioni provvisoriamente rifiutate possano andare "a testo", e viceversa: perché non permettere di farlo? Non è forse l'edizione, secondo il magistero di Contini, l'ipotesi più economica che dà conto dei dati disponibili? Ma se è così, allora essa deve e può cambiare a fronte dell'evoluzione dei dati testuali, all'emergere di nuovi dati, o semplicemente in conseguenza di una migliore e più perspicua interpretazione di dati già noti.

4) Alla nettissima gerarchia di importanza scientifica fra l'edizione critica propriamente detta e l'edizione diplomatica, o diplomatico-interpretativa, si oppone ora la stessa centralità del concetto di codifica; se l'edizione informatica ruota attorno a una codifica, ecco allora che diventa del tutto vero il paradosso enunciato tanto tempo fa dal paleografo Masai, cioè che solo le edizioni

diplomatiche sono definitive, e che sulla loro base si potranno poi costruire *n* edizioni critiche, tutte aggiornabili o revocabili in dubbio.

5) Infine, alla esclusiva autorialità dell'editore, cioè al carattere assolutamente individuale e non collettivo del lavoro di edizione, si oppongono semplicemente la realtà e il buon senso. Se quanto detto sopra rispetto al punto 1 è vero, ecco allora che si può immaginare un'edizione a cui collaborino più editori, sia nel senso più stretto e diretto (che ad essa possono apportare pareri e proposte diversi editori online) sia nel senso più vasto, ma altrettanto importante teoricamente, che ad essa vengono *linkate* altre edizioni disponibili, o edizioni diverse dello stesso testo, oppure l'intertesto o l'intratesto a cui il testo da editare rinvia.

6. Dunque forse dobbiamo cominciare a pensare (per entrare davvero nella "seconda fase" della filologia informatica) a: 1) un'edizione sempre *in fieri*, mai definitiva (esattamente come è nella natura della ricerca, di ogni ricerca); 2) un'edizione che non ricostruisce necessariamente una e una sola versione (o stato) del testo, ma che può ripercorrerne e riprodurne (con la "rigorosa mobilità" di cui è capace la macchina informatica, del tutto impossibile alla stampa!) *il movimento* del testo nella tradizione; 3) un'edizione che lascia aperto il vitale rapporto fra testo e apparato, o fra diversi rami della tradizione, o fra diverse versioni varianti, ecc.; 4) un'edizione che concentra il peso del lavoro scientifico sulla codifica (ma dunque – per dir così - sulla forma moderna e informatica dell'edizione diplomatica); 5) infine un'edizione che ha più padri, o madri, e che non appartiene in esclusiva a nessuno, perché è frutto di una sorta di cervello collettivo, della comunità scientifica, della "comunità degli interpreti".

Oltre i confini del testo: le Digital Humanities tra scienza e opportunità

Federica Perazzini

Nel tentativo di definire in modo preciso ed esaustivo la disciplina oggetto del nostro convegno oggi, credo che nulla si rivelerà più efficace e paradossalmente calzante dell'aforisma coniato giusto qualche giorno fa da Simon Mahony, condirettore del Centre for Digital Humanities presso University College London: «Digital Humanities is that thing that we, Digital Humanities people, do». Ciò che sembra emergere da questa simpatica tautologia è, in realtà, un genuino disagio, se non direttamente l'impossibilità, di dipanare il mistero di cosa effettivamente si nasconda dietro l'ammaliante etichetta "Digital Humanities"; una difficoltà che io stessa mi sono trovata ad affrontare molteplici volte nel corso degli ultimi due anni e che questa relazione vuole oggi esplorare e, qualora possibile, superare.

In ciò che segue, infatti, cercherò di esaminare cosa significa, a quasi settant'anni dalle prime sperimentazioni, essere nelle *digital humanities*, quali interessi e quali opportunità, anche di tipo

economico, si nascondono dietro l'evoluzione di questa disciplina e soprattutto, con un occhio alle nuove generazioni, su quali possibili terreni di sfida potrà delinearsi il suo futuro.

Per fare questo, presenterò brevi cenni sulla storia della disciplina dimostrando come, di fatto, essa possa rivendicare il ruolo di nuova scienza in questa nostra era di rivoluzione digitale. Verranno inoltre illustrati degli esempi concreti di progetti di ricerca concernenti l'analisi computazionale di corpora sottolineando fino a che punto questa metodologia possa giovare al campo degli studi umanistici apportando della nuova conoscenza.

Verso una definizione-indefinita

Per rispondere al primo quesito proposto, ossia come poter definire oggi le Digital Humanities, comincerò col riportare la spiegazione che ne dà l'*Oxford University Computing Services Guide to Digital Resources for the Humanities* (2001): «digital technology in studying traditional humanities objects»; magie computazionali applicate alle scienze umanistiche. Una definizione senz'altro corretta, ma non del tutto esaustiva. Se è vero, infatti, che fin dalle origini questa disciplina si è sempre configurata come una sorta di metodologia ausiliaria utilizzata a potenziamento della capacità d'analisi dei singoli ricercatori, oggi le *digital humanities* sono andate oltre questa logica strumentale proprio a causa dell'innegabile rivoluzione che, da circa una decina d'anni, ha cambiato il nostro modo di conservare e concepire l'informazione digitale. Un periodo spesso definito "The Age of the Big Data", un'epoca in cui più di

ventiquattro secoli di storia culturale umana in termini di manoscritti, testi stampati, immagini e artefatti sono stati trasformati in oggetti digitali, accessibili e "laborabili": una miniera d'informazione virtuale da più di 500 miliardi di gigabyte che attende solo di essere scavata¹.

Nel solo ambito letterario, compagnie private come la ProQuest sono entrate nel business delle Digital Humanities procedendo alla scannerizzazione, digitalizzazione e archiviazione di milioni di libri, principalmente in lingua inglese, creando dei corpora specializzati e organizzati secondo un criterio cronologico. Tra questi ricordiamo ECCO, archivio della letteratura canonica e non canonica del diciottesimo secolo o la Chadwick Healey collection, composta da classici e best sellers del diciannovesimo secolo. La stessa British Library poi, nel febbraio del 2010, ha dichiarato di essere pronta al rilascio di 65.000 prime edizioni di romanzi e altre opere non funzionali risalenti al diciannovesimo secolo in formato Kindle. Accanto a queste raccolte private, per altro vendute dietro un corrispettivo di centinaia di migliaia di dollari, coesistono poi biblioteche digitali ad accesso aperto e gratuito come l' "Internet Archive", col suo catalogo di 400 autori e 40 lingue, o l'ormai famosissimo "Google books", un progetto che dal dicembre 2004 monopolizza il panorama della digitalizzazione internazionale con

¹ Proprio la metafora della miniera viene evocata nel termine Americano "data mining" utilizzato per descrivere la metodologia d'investigazione delle *digital humanities*.

più di 5 miliardi di testi della più svariata natura disponibili in formato OCR (Optical Character Recognition).

Tenendo questi dati a mente, credo sia dunque piuttosto riduttivo continuare a pensare alle *digital humanities* come semplice mezzo tecnologico prestato a fini letterari e, nel tentativo di una più consona ridefinizione, mi sono rivolta alla *Metodologia delle scienze sociali* di Max Weber. In uno dei saggi che compone questa raccolta, *Oggettività nelle scienze e nelle politiche sociali*, lo studioso tedesco fornisce il perfetto substrato teorico per inquadrare il ruolo di questa disciplina nell'odierno panorama di rivoluzione digitale quando afferma che una nuova "scienza" emerge quando un nuovo problema viene approcciato con un nuovo metodo². Proprio in virtù di questo semplice ma efficace assioma, è possibile riconoscere alle *digital humanities* lo status di nuova "scienza" dell'era dei grandi dati; un metodo che trova la sua giustificazione e le sue stesse condizioni di possibilità nelle nuove problematiche della testualità digitale che questa nostra era post-google books comporta.

Ma può bastare questo semplice sostegno teorico per far sì che le *digital humanities* vincano l'ormai annosa diffidenza da parte dell'establishment accademico e vedano finalmente riconosciuto il proprio ruolo di disciplina autonoma e protagonista della ricerca umanistica contemporanea? Temo la risposta sia no e temo, altresì,

² M. Weber, "'Objectivity' in Social Science and Social Policy". In E.A. Shils, and H.A. Finch (Ed.), *The Methodology of the Social Sciences. Max Weber*. New York, The Free Press. 1969/1904, pp. 50-112.

che sia ancora valida la teoria di Jerome McGann, fondatore dello SPECLAB alla Virginia University, per cui per essere prese sul serio dall'accademia con la "A" maiuscola, le *digital humanities* debbano ancora dimostrare sul campo la loro capacità di apportare un plusvalore nell'esistente scenario degli studi letterari:

[...] the general field of humanities education and scholarship will not take the use of digital technology seriously until one demonstrates how its tools improve the ways we explore and explain aesthetic works – until, that is, they expand our interpretational procedures³.

Dal qualitativo al quantitativo, per una nuova conoscenza possibile

Per dimostrare quanto le *digital humanities* possano contribuire in modo inedito ed esclusivo all'avanzamento degli studi culturali, credo sia opportuno iniziare a sfruttare al meglio le potenzialità connaturate al mezzo informatico e abbracciare più ampie prospettive critiche. Nello specifico, credo sia utile operare un cambio di paradigma e spostarsi dall'usuale piano analitico-qualitativo, in cui ci si concentra nel dettaglio sulle singole parti componenti un certo fenomeno, ad un piano sintetico-quantitativo, in cui il tutto è maggiore della forma delle sue parti e l'oggetto di studio viene inteso ed esaminato nel suo insieme. A livello di

³ J. McGann, *Radiant Textuality: Literary Studies After the World Wide Web: Preface*, New York, Palgrave MacMillan, 2004/2001, p.xii.

approccio analitico-qualitativo, le *digital humanities* vantano una lunga tradizione di esperimenti di *close-readings* potenziata e innovativa; due esempi per tutti di questo metodo sono i grandi studi monografici di Jonathan Hope sull'*authorship* del corpus shakespeariano⁴ o quello di J. F. Burrows sui romanzi della Austen⁵.

Al contrario, una visione sintetica-quantitativa mirante all'individuazione di macro-tendenze all'interno di ampi corpora digitali è ciò che anima gli esperimenti di macro-analisi, per l'appunto, computazionale del Literary Lab della Stanford University o il recentissimo *Culturomics*⁶ di Harvard. In questa ultima sezione, presenterò due esempi di progetti di ricerca realizzati attraverso la macro-analisi computazionale a cui ho avuto modo di lavorare durante il mio periodo di *visiting* allo Stanford Literary Lab. Il fine è quello di illustrare quali vantaggi questo tipo di ricerca

⁴ J. Hope, *The authorship of Shakespeare's plays: a sociolinguistic study*, New York, Cambridge University Press, 1994/1962.

⁵ J. F. Burrows, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*, Oxford, Clarendon Press, 1987.

⁶ Questo neologismo è stato recentemente introdotto nell'articolo *Quantitative Analysis of Culture Using Millions of Digitized Books*, pubblicato su «Science» nel dicembre 2010 dai ricercatori di Harvard Jean-Baptiste Michel and Erez Lieberman Aiden. In base a quanto emerso dal suddetto articolo, *Culturomics* è una forma di lessicologia computazionale che studia il comportamento umano e i fenomeni culturali attraverso l'analisi quantitativa di testi digitali. L'idea è che tali fenomeni si riflettano nel linguaggio e nell'uso delle parole e che quindi, studiando la tendenza e la frequenza di certi lemmi, sia possibile prevederli. J. B. Michel. E. Lieberman Aiden, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, et al., "Quantitative Analysis of Culture Using Millions of Digitized Books". In *Science*, 331 (6014), 16 December 2010, pp. 176–82.

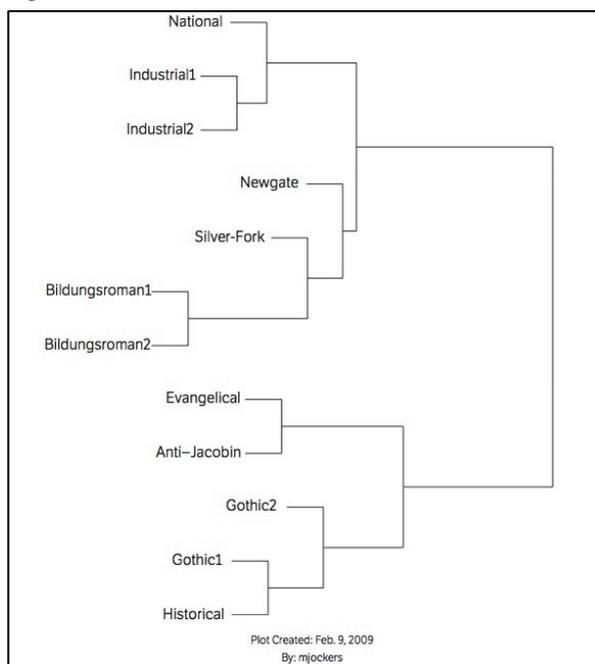
comporta e in che modo essa può dimostrarsi in grado di apportare nuova conoscenza nel campo delle scienze umane.

Due Esperimenti: MFW e Campi Semantici

Quantitative Formalism è il titolo del lungo esperimento che ha visto impegnati, dall'autunno 2008 a marzo 2009, la squadra dello Stanford Literary Lab di Franco Moretti, e Michael Witmore dell'University of Wisconsin. In questo pionieristico lavoro, pubblicato poi nel gennaio 2011 come on-line pamphlet⁷, si è testata l'ipotesi di una procedura informatica *unsupervised* per la separazione ed il riconoscimento dei sottogeneri romanzeschi attraverso le MFW (most frequent words). Il campione di riferimento era composto da 48 romanzi appartenenti a 12 generi differenti. Ognuno dei testi campione poteva considerarsi un ben acclamato classico del genere d'appartenenza (per esempio, *The Monk* per il gotico, *Daniel Deronda* per il romanzo di formazione o *Hard Times* per il romanzo industriale).

⁷ litlab.stanford.edu/LiteraryLabPamphlet1.pdf

Figura 1.



Con grande sorpresa degli addetti ai lavori, il computer riuscì ad operare una "clusterizzazione" automatica perfetta. Perfetta in quanto totalmente coincidente alla classificazione della critica letteraria tradizionale da cui si era partiti. Ulteriore fonte di stupore, però, riguardava "come" la macchina fosse giunta al medesimo risultato dell'uomo. Infatti, nel commentare il brano a più alto coefficiente "gotico" selezionato dal computer, Sarah Allison, co-autrice del pamphlet, osserva:

[...] the gothic of Docuscope was different from that of "Humanscope" (as she called it): it was not the same gothic we saw. For us, that page was gothic because of the subdued terror and the archway, the ruin and apprehension and the limbs that trembled – not because of the "he" "his" "him" "had" "was" "struck the" and "heard the" which caught Docuscope's attention.⁸

"Docuscope", il programma di MFW utilizzato, era stato in grado di classificare come appartenente al genere gotico un brano tratto da *A Sicilian Romance* (1790) della Radcliffe – romanzo gotico per antonomasia - ma ci era arrivato attraverso degli indicatori (articoli, pronomi e tempi verbali) totalmente diversi da quelli abitualmente considerati dall'uomo (elementi tematici come terrore, passaggi sotterranei ed eroine tremanti).

In conclusione, questo esperimento aveva dimostrato essenzialmente due cose. La prima riguardava senz'altro la capacità da parte della macchina di "riconoscere" e quindi raggruppare romanzi in base al genere senza alcun intervento umano. La seconda, conseguente alla prima e dalla portata del tutto inaspettata, riguardava il fatto di aver portato alla luce l'esistenza di tutto un substrato di micro-unità formali quali articoli, prefissi, pronomi e preposizioni che, al pari di più evidenti aspetti semantici, era in grado di fornire un netto segnale di genere. Per dirla con una metafora biologica, questo esperimento era riuscito ad isolare i

⁸ S. Allison, R. Heuser, M. Jockers, F. Moretti, M. Witmore, *Quantitative Formalism: An Experiment*. Stanford on-line pamphlet, 2011, p. 8. <http://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>.

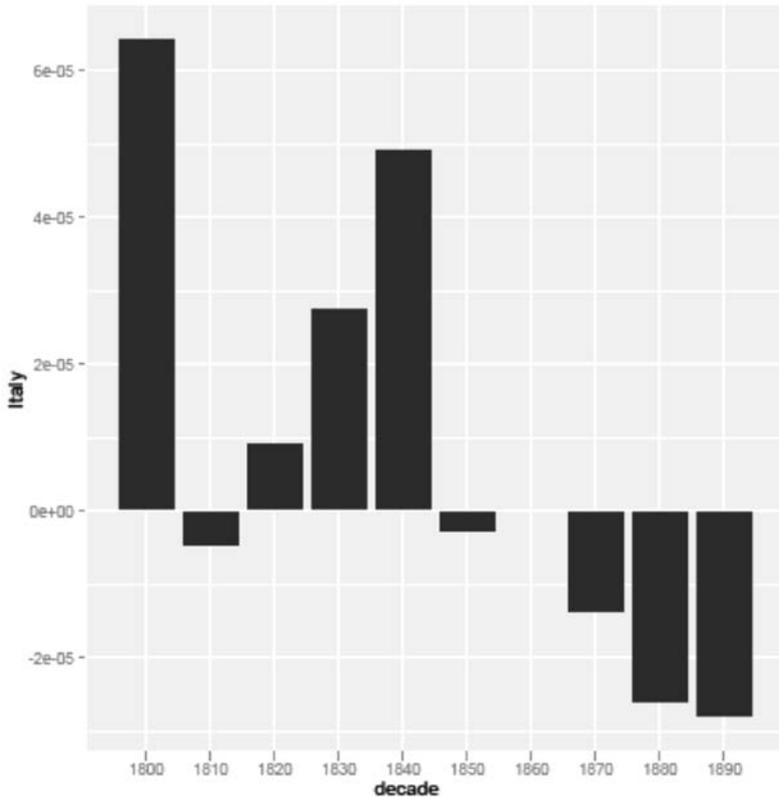
singoli geni che compongono il DNA lessicale dei diversi generi letterari del romanzo inglese.

Il secondo esperimento dello Stanford Literary Lab, invece, riguardava l'applicazione di un generatore di campi semantici empirici allo studio di luoghi geografici sempre all'interno del romanzo inglese dell'Ottocento. Nello specifico, il generatore di campi semantici, detto "Correlator", è uno script che permette di correlare ad una parola chiave inserita tutti gli altri lemmi che, all'interno del corpus di riferimento, condividono la sua stessa frequenza d'uso e linea di tendenza cronologica. In questo modo, pur non trattandosi di un campo semantico in senso tradizionale, il "correlator" può comunque definirsi un generatore di campi semantici di tipo empirico proprio per questa sua capacità di mettere in relazione parole in base ad un criterio di contiguità e probabilità di co-occorrenza.

Lo scopo di questo secondo esperimento era dunque quello di tracciare una mappatura dei luoghi del romanzo dell'Ottocento inglese utilizzando il "correlator" per investigare sull'immaginario simbolico connaturato a ciascuno di essi. A tale scopo, l'intero corpus di romanzi dello Stanford Literary Lab era stato previamente processato con NER, un'applicazione di "entity recognition" in grado di estrarre e classificare singole unità testuali in base a categorie predefinite quali nomi propri, quantità o percentuali. Nel caso della nostra ricerca, il criterio di selezione era appunto quello geografico e alla fine della procedura ognuno dei 3.603 romanzi del database era stato dotato di tag "location". Successivamente, usando uno script di PHP, tali dati sono stati visualizzati in una matrice in cui, ad ogni titolo del corpus, veniva accostato l'esatto numero di occorrenze di

ciascuna location. Una volta proceduto con i dovuti calcoli statistici per normalizzare i risultati, sono stati prodotti dei grafici espositivi attraverso il celebre programma 'R'⁹.

Figura. 2



⁹ The 'R' Project for Statistical Computing (www.r-project.org).

Interessante è il caso dell'Italia che, presente in tre diversi set di location del nord (Milan, Venice) del centro (Rome, Florence, Tuscany) e del sud (Naples, Palermo, Sicily), appare con regolarità nelle pagine degli scrittori britannici fino alla fine del 1860 ed è caratterizzata da sfumature semantiche diverse per ciascun'area e ascrivibili a diversi generi letterari.

Tabella 1.

Keyword: SICILY		
artifices (+0.89130)	hero's (+0.84810)	informing (+0.82763)
dungeons (+0.88229)	fails (+0.84545)	philanthropy (+0.82424)
assiduously (+0.88071)	unmindful (+0.84443)	horrors (+0.82384)
inquisitor (+0.87937)	monk (+0.84399)	medicines (+0.82373)
dreadful (+0.87777)	mortally (+0.84201)	endeavor (+0.82347)
ignominy (+0.87315)	horror (+0.84192)	hero (+0.82274)
recurred (+0.87191)	grieves (+0.84136)	misfortunes (+0.82245)
signora (+0.87131)	atrocious (+0.84032)	don (+0.82160)
convent (+0.86628)	obdurate (+0.83827)	deluded (+0.82148)
trembled (+0.86267)	dungeon (+0.83539)	vicious (+0.81727)
assassin (+0.86151)	confident (+0.83515)	emotions (+0.81552)
fear (+0.85964)	perfections (+0.83344)	embracing (+0.81379)
Madrid (+0.85547)	amiable (+0.83287)	conjectures (+0.81296)
court (+0.85539)	fainted (+0.83246)	alleviate (+0.81244)
tormented (+0.85472)	caprices (+0.83169)	artful (+0.81220)
gaming (+0.85406)	extremely (+0.82924)	nun (+0.81211)
rigid (+0.84978)	rural (+0.82767)	mournfully (+0.81141)

Tabella. 2

Keyword: TUSCANY		
staggering (+0.96213)	bust (+0.90725)	sprang (+0.89204)
purchases (+0.95717)	shirt (+0.90638)	appreciate (+0.89170)
flinging (+0.95351)	blanched (+0.90628)	awaiting (+0.89094)
grinning (+0.94502)	drawing (+0.90564)	handkerchiefs(+0.8894)
glance (+0.93277)	tone (+0.90559)	expression (+0.88300)
thrusting (+0.93115)	slap (+0.90467)	wink (+0.88287)
confidential (+0.92511)	cushion (+0.90395)	pavement (+0.88171)
hissing (+0.92213)	overhanging (+0.90366)	clothing (+0.88157)
winked (+0.92143)	interposed (+0.90353)	glass (+0.88056)
exciting (+0.91920)	glancing (+0.90122)	tranquilly (+0.88019)
aloud (+0.91736)	astounding (+0.90119)	table (+0.87884)
spinster (+0.91728)	unusual (+0.89982)	boudoir (+0.87869)
glimpse (+0.91629)	tumbler (+0.89636)	oriental (+0.87108)
intently (+0.91537)	bolted (+0.89588)	waistcoat (+0.87058)
ghastly (+0.91229)	handkerchief (+0.89500)	mother-in-law(+0.86992)
sternly (+0.91215)	rambling (+0.89443)	hand(+0.86924)
moment's (+0.91183)	satin (+0.89236)	mastery
pause (+0.90799)	loan (+0.89236)	

Le cifre tra parentesi corrispondono al valore di deviazione standard con cui ciascuno dei lemmi generati dalla ricerca si correla alla parola chiave inserita; più il valore si avvicina ad 1 più le due parole risulteranno avere una frequenza d'uso analoga e quindi, per estensione, saranno in relazione tra loro a livello empirico. Nella fattispecie, la parola "artifices", artifici, è correlata alla parola sorgente "Sicily" con un punteggio di +0.89130 di deviazione standard sopra la media del corpus. "Purchases", acquisti, è correlata alla parola chiave "Tuscany" con +0.95717 di deviazione standard sopra la media del corpus e così via.

Il contributo di questo esperimento sta nell'aver provato quanto all'emergere di nuovi spazi geografici all'interno dell'immaginario del romanzo inglese corrisponda la nascita di uno specifico genere letterario. Infatti, se all'inizio del diciannovesimo secolo il picco delle *locations* italiane del sud (Palermo, Naples, Sicily) è chiaramente connesso allo sviluppo dell'immaginario del genere gotico (con nette correlazioni semantiche come "prigioni", "inquisitori" e "terrore"), negli anni 40' e 50' dell'Ottocento, il sorgere di nuove aree quali "Rome", "Florence" e "Tuscany" comporta un cambiamento dell'identità simbolica italiana verso "cuscini", "satin" e "prestiti" più in relazione con i temi di nobiltà decaduta del *silver fork novel*.

Come d'abitudine, alla fine di ogni riunione dello Stanford Literary Lab, qualcuno dei partecipanti deve vestire i panni dell'avvocato del diavolo e porre al gruppo l'odiosa questione: "Is this something we didn't know?". Un interrogativo che tradisce l'ansia e la pressione di questi giovani ricercatori non solo nel dover

trovare con i loro esperimenti qualcosa di nuovo, ma anche nel dover imparare a controbattere le critiche sempre aspre e provenienti da ogni direzione. Ebbene per quello che ho visto, la risposta è comunque sì; sì è stato detto qualcosa di nuovo; sì, è stato portato alla luce qualcosa che non sapevamo e che, più importante ancora, non avremmo mai potuto scoprire con la ricerca tradizionale. Ma, soprattutto, ciò che veramente costituisce il valore aggiunto di questo tipo di ricerca è che ogni esperimento, se eseguito in correttezza, contribuisce alla creazione di una nuova scienza delle forme letterarie in cui lo studio di fatti artistici e fenomeni estetici può e deve essere finalmente verificabile a livello oggettivo.

Conclusioni: prospettive e suggerimenti

Ho cominciato questo percorso ponendomi l'obiettivo di investigare cosa significasse oggi essere nelle *digital humanities*. Definito l'oggetto di studio ed esplorato a livello teorico e pratico le potenzialità della metodologia, a questo punto non ci resta che affrontare l'ultimo dei quesiti posti nell'introduzione, delineando gli ipotetici scenari futuri di questo grande work-in-progress e ponendo particolare attenzione alle trasformazioni che si imporranno alla figura del critico.

Dovendo fare una previsione, infatti, credo che la grande partita si giocherà nell'ambito dei progetti di ricerca. Essere preparati ad investire in sperimentazioni e analisi dagli esiti non sempre sicuri, evitando di cedere alle facili lusinghe di un'applicazione della disciplina al solo campo dell'archiviazione e della conservazione,

sarà fondamentale. Se è vero, infatti, che biblioteche virtuali e altre forme di custodia del patrimonio in formato digitale sono progetti piuttosto proficui, e non solo a livello intellettuale, è altrettanto vero che questi si rivelerebbero uno spreco in termini di opportunità qualora i dati non potessero essere esplorati e studiati.

Un'altra cosa a cui bisognerà essere preparati, e che forse si rivelerà la prova più complessa da superare, è il cambiamento della figura del critico-ricercatore che questo tipo di metodologia d'investigazione richiede. Infatti, nell'ottica di una efficace sperimentazione computazionale, sia essa relativa all'analisi del testo o di altre forme culturali, tre sono le caratteristiche che lo studioso dovrebbe sviluppare: autonomia gestionale, capacità di lavorare in gruppo e cultura del fallimento costruttivo.

Per autonomia gestionale s'intende la capacità da parte del ricercatore di interfacciarsi a livello elementare ma indipendente con gli strumenti e applicazioni informatiche necessarie al suo lavoro. Per fare un esempio pratico, questo potrebbe tradursi nella possibilità d'interrogare in tutta autonomia un database di testi, generare una lista di frequenza, normalizzare dei risultati con una formula di deviazione standard ecc. In assenza di una tale conoscenza tecnica del mezzo informatico, tra l'altro non richiesta ad uno studioso di scienze umane, il ricercatore dovrà delegare del personale tecnico a procedere all'estrazione dei dati per suo conto. In questo scenario, l'autonomia gestionale corrisponderebbe al raggiungimento di una certa preparazione di base da parte del ricercatore per dialogare efficacemente con i tecnici ed instaurare

una collaborazione più fruttuosa. Non è necessario, infatti, né tantomeno pratico trasformare ogni filologo o critico letterario in un ingegnere informatico; ciò che è auspicabile, invece, è che gli umanisti imparino che tipo di domande rivolgere ai colleghi informatici e che tipo di risultati, in termini di dati, potersi aspettare da loro.

Per ciò che riguarda la capacità di lavorare in gruppo, invece, sebbene questo non sia un concetto particolarmente oscuro a livello di significato, è importante spiegare i motivi per cui la ricerca computazionale sembra imporre una modalità di lavoro decisamente più collettiva che individuale. Innanzi tutto, c'è un discorso di scissione tra la fase di creazione dei dati e il momento della loro interpretazione. Come si è visto in riferimento all'autonomia gestionale, infatti, nella ricerca computazionale è più che consueto che il processo di generazione dei dati venga completato da una persona diversa da colui che dovrà procedere con la loro analisi. Una tale separazione, oltre che necessaria a fini tecnici, è ulteriormente positiva proprio in quanto garantirebbe una migliore oggettività dell'analisi allontanando il ricercatore dalla possibilità e dalla tentazione di manipolare le fonti a sua disposizione, in questo caso i dati, e usarle a giustificazioni di proprie teorie o ipotesi precostituite. Inoltre, per quello che riguarda la fase interpretativa, la quantità e la varietà di dati estratti a cui fare riferimento può raggiungere dei volumi talmente ampi da richiedere un vero e proprio sforzo collettivo per riuscire ad organizzarli ed analizzarli tutti.

Infine, il fallimento. Una delle primissime cose che Matthew Jockers, co-fondatore dello Stanford Literary Lab, mi disse una volta

terminato di illustrargli il mio progetto di tesi, una macro-analisi computazionale del romanzo gotico, fu: "*Remember that failure is an option*". Il fallimento è un'opzione. Lì per lì non capii né l'intento né il significato di quelle parole di cui rimasi molto dispiaciuta; fortunatamente oggi, a distanza di un anno, esse hanno per me tutto un altro suono. Non solo il fallimento è un'opzione, è un motore; l'essenza stessa della ricerca scientifica. Non so se il prof. Jockers pensasse davvero che la mia tesi fosse un fiasco, so solo che con quelle parole mi ha rivelato il vero significato ed il vero piacere di fare ricerca; due cose che in questi tempi di crisi e di incessanti attacchi al mondo dell'istruzione e della cultura credo sia bene continuare a tenere a mente.

Digital editions between embedded markup and external representation. A case study: Vespasiano da Bisticci's Letters

Francesca Tomasi

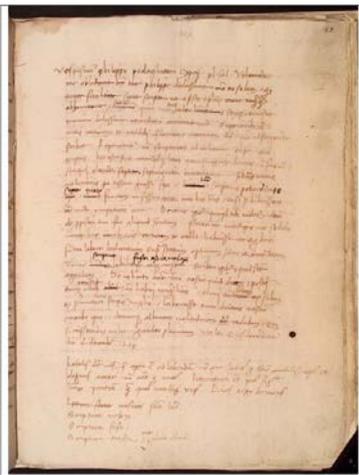
The case study

Creating a digital edition means first creating an adequate markup model. The encoding scheme must respond both to the nature of the primary source and to the computational objectives; it's necessary to consider, on the one hand, which text features we want to focus on, and on the other, what kind of digital representation we want to give them. This means reasoning about data, with the aim of creating suitable information and extracting knowledge from the source.

The case study presented here is a collection of letters in Latin and XV century Italian, now held in different libraries and, mostly, archives. The correspondence documents the personal and professional relationships managed by the Florentine librarian and copyist Vespasiano da Bisticci, who was also the leader of a school of copyists, maker of some manuscript repositories of different European libraries.

The correspondence is with notable people of that period and the content regards mostly the trade of manuscripts copied, proposed or requested by/to Vespasiano. A lot of these manuscripts copied by Vespasiano's school had been identified in codices now held in various libraries. From the letters it's possible to learn about features of these realized manuscripts: the materials, the copyists, the costs, but also the names of the Latin, Greek and Humanistic authors and texts that were the most fashionable at the time.

Figura 1.

	Incipit	Vespasiano Philippo Podaghatero Cyprina diurimum salutem
	Body	<p>Vehementer me oblectant littere tue, Philippus dulcissime, non eo solum quod tanta facilitate sunt scripte, ut a prisco epistolarum modo nequaquam abhorreant, sed etiam quia cum eas video semper recusat in animo dulcissima recordatio amicitie nostrae.</p> <p>Superioribus vero meis certiorum te reddidi Florentie neminem esse qui ad fragmenta scribat. Reperirentur vero scriptores ad volumina eo pacto quo exoptas, hoc est ut unumquodque latus quinquaginta lineas, versus vero singuli elementa septuaginta continerent.</p> <p>Pretium unius voluminis essent grossi sex. Littere scriptoris potiores tuis sunt, quarum formam misissem certe cum his litteris, nisi prohibuissent nonnullae occupationes meae. Decerne igitur quicquid tibi videtur, deinde de proposito tuo fac aliquid sentiam.</p> <p>Praeterea, ut intelligas me sedulo invigilare rebus tuis, certiorum te reddo habuisse me eisdem summo labore Lactantium super Statium, optimum sane et emendatum. Forma scripturae fusa est et velox. Scribe igitur quid sit agendum.</p> <p>De Roberto autem nostro nescio quid dicam: postquam enim recessit a nobis, non habui unquam litteras. Novi solum ex Procuratore regis Anglie laborasse eum diutius nescio morbo quo, demumque ad bonam validudinem esse redactum. Qui si impresentiarum valet, gaudeo plurimum.</p>
	Explicit	Vale Is Florentia, die 7 decembris 1545

<p>SENDER and ADDRESSEE</p> <p>SALUTATIO (in incipit)</p>	<p>TECHNICAL LEXICON</p>	<p>MENTIONED MANUSCRIPTS</p>	<p>NAMES (PROPER NAMES AND VARIANT FORMS)</p>	<p>SALUTATIO (in explicit)</p> <p>DATA (place and date)</p>
---	---------------------------------	-------------------------------------	--	---

Here an example of a letter by Vespasiano da Bisticci to the humanist Filippo Podocataro (Firenze, 4 December 1448). In the letter Vespasiano spoke about the production of a *Lactantium super Statium* for Filippo and describe the features of the codex realized.

The element of markup are emphasized with colors, in order to define the different levels of representation of the different phenomena: sender and addresses, *salutatio formulae*, technical lexicon, mentioned manuscripts, persons, datatio (place and date).

The letter is a page of a manuscript held in Firenze, Bibl. Naz., shelfmark: Magl. VIII, 1390, f. 102.

The purpose of the digital edition is on one hand to represent the information that is implicitly connected inside the source (people with manuscripts, manuscript with lexicon); on the other hand to create semantic links with WWW resources, useful in order to describe this information (people, manuscripts, lexicon).

The first purpose is thus to create relationships between people and manuscripts related to people at some level (the copyist, the owner, the requester) and the manuscripts with the lexicon used in order to describe them; this information born naturally from the letters and have to be expressed in a formal way. The second aim is to create relationships between people and resources useful to describe people (public prosopography), mentioned manuscripts with the existent codices (catalogues of manuscripts), the lexicon with repositories of technical words (thesauri).

The traditional apparatus of scholarly editions have, in this way, to redefine its tasks and change its field of interest and relevance. It documents the information that the editor is able to recognize from reading the texts (implicit relationships), but also

from analyzing those related resources, able to focus on the relevant elements in the source. The dialogue between annotated elements is opened to the web of resources.¹

The embedded markup

The markup model presented covers all the different aspects of the edition, that is the model reflects the different access points to the letters content. This means that using the XML embedded markup it is possible to automatically extract indices of all the annotated elements (i.e. with XSL(T)). The entries of such indices are the possible paths of users' navigation within the letters content ("faceted" navigation in a Web perspective).

The focus of such indices is on:

- a) persons mentioned in the letters;
- b) texts cited and manuscripts realized by Vespasiano's school (regarding Latin, Greek and Humanistic texts and authors);
- c) the technical lexicon of the copy and of librarian trade.

At the markup level it is quite easy to represent, with the appropriate TEI elements, all this information. Here some examples of annotation (in TEI P5) of all the above-mentioned text features, with external references to RDF files (described in section 3).

¹ The decision to use the embedded XML/TEI markup, for the first step of annotation, derived from two strictly connected considerations: the first is that the majority of digital editions are marked up in this way and so it's necessary to start from a shared situation, in order to give a knowledge representation strategy; the second is that the project was born in 2000 when the XML/TEI markup was one of the most used annotation methods.

a) <persname> and <rs>

Identify mentioned persons with a sigla

Raccomandami a

```
<persname ref="people.rdf#PM">Piero</persname> et a  
<persname ref="people.rdf#DA">Donato</persname> et alla brigata  
di casa, et a messer  
<persname ref="people.rdf#GA">Giovanni Argiropolo</persname>  
et a tutta l'Accademia sopra tutti  
<persname ref="people.rdf#LM">Lorenzo</persname>.
```

Identify persons named in different way with the same sigla

```
<persname ref="people.rdf#AC">Alfonso duca di Calabria  
</persname>
```

[...] Quanto alla parte dello Illustrissimo

```
<rs type="person" ref="people.rdf#AC">nostro Primogenito</rs>,  
havemo gran piacere habia satisfacto alli animi de quisti excelsi  
Signori et homini de questa città, et molto li restamo obligati  
dello honore che per quelli li hè stato facto allo prefato  
<rs type="person" ref="people.rdf#AC">Duca</rs>.
```

b) <bibl> with <author> and <title>

A mentioned manuscript (Plutarchus, *Vitae*)

```
<bibl ref="manuscripts.rdf#PV_1">  
  <title>Vite</title> di <author>Plutarco</author>  
</bibl>
```

c) <term>

Some technical terms

<term type="writing" ref="lexicon.rdf#aF">ad fragmenta</term>

<term type="support" ref="lexicon.rdf#iP">in papiro</term>

<term type="binding" ref="lexicon.rdf#C">chordovani</term>

<term type="illumination" ref="lexicon.rdf#M">miniare</term>

The base TEI markup let us identify: proper name (<persname> and <placename>) and referring string (<rs> with @type for specification), mentioned manuscripts (<bibl>) and related author (<author>) and title (<title>), technical term (<term>) on various field (@type).

The @ref value in the instances of persname, placename and rs allows a first identification of individuals being mentioned: missing parts of names are solved outside (e.g., the person being identified simply as 'Piero') and variants are associated to the same instance (e.g., the 'nostro Primogenito' and the 'Duca', all referring to the same 'Alfonso, duca di Calabria').

The same attribute @ref was used also for the element bibl and term, with a TEI customization. The @ref value points to the description of the items, store in a place outside the document. Three .rdf files was created, for the three mentioned categories: one for people (people.rdf), one for manuscripts (manuscripts.rdf) and another for the lexicon (lexicon.rdf). Therefore, within these elements, an access point is defined as an element owning a URI reference that point to an @rdf:about attribute in the external repository. Using the URI mechanism [see also 3] it is possible to create a connection between the content of the letters and the information represented and stored outside. The value of the

attributes @ref in the examples above, therefore, are URIs to the items of the store where this information can be formulated in a formal and structured way (i.e. "people.rdf#PM" points to the section of the RDF file where Piero de' Medici, called with the "PM" sigla inside the document, is described).

If the URI pointing give the possibility to formally describe annotated elements in the external representation, it's necessary to define a system to create connections between these elements. These connections must answer to questions like: which relation exists between 'Piero de' Medici' ("people.rdf#PM") and one specific exemplar of Plutarchus's *Vite* ("manuscript.rdf#PV_1")? And which one between the same manuscript ("manuscript.rdf#PV_1") and the word 'chordovani' ("lexicon.rdf#C")? But the model has to answer also to questions like: which exemplar of Plutarco's *Vite* has been realized and is it available in any European library? What the word 'ad fragmenta' means? Who 'Piero de' Medici' is?

For the first set of questions some more formal semantics is needed, in order to describe relationships between annotated data elements. Add semantics to a document doesn't mean to add specific attribute and relative values but means to find a way to express in a formal way a machine-processable information. RDF gave the possibility to express asserts on the different properties (external information) related to each category (people, manuscript, lexicon).

For the second set of questions it's necessary to use techniques useful in order to connect the edition with existent resources on the WWW, able to extend the description of the annotated elements, that is to enable the dialogue between our URIs and other URIS referred to the same concept.

The external information

Once the texts are annotated it's necessary firstly to focus on what kind of additional information it is possible to define externally and to make it accessible via URI-based pointing in the RDF files. In this way it's possible to create a public authority list, accessible by researchers via PSI (*Public Subject Indicator*), like in the EATS system²

Starting from Vespasiano's letters it became possible to expose authority information about people, manuscripts, technical lexicon regarding XV century culture, in order to start to create a public authority list to be integrated in authority records.

We need firstly to decide which kind of relevant added data it is possible to specify for each of the three categories. Then we have to reflect on which kind of relationships we could define.

a) Persons = people.rdf <rdf:description rdf:about="value-of-related-URI">

At this level a series of information (or properties) could be given regarding a person:

- PSI (extracted from URI);
- normalization of the name (e.g., forename + surname. Entry for the authority list);
- variant forms of the name (automatically extracted by the same URIs in the documents);
- role (e.g., count, prince, duke, king, bishop, cleric, humanist, copyist);

² J. Norrish, *EATS: an Entity Authority Tool Set*. NZETC (New Zealand Electronic Text Centre), 2007. URI: <http://hdl.handle.net/10063/220>. Database Access: <http://authority.nzetc.org>.

- biography:
 - created when not already existent;
 - written by Vespasiano himself in his, *Vite degli uomini illustri* (digital versions);
 - present in (one or more) other repertories – link via URI if possible.

- **relationships:**
 - with other persons – i.e. brother-of, son-of, father-of;
 - with places – i.e. king-of, born-in, died-in;
 - with dates – i.e. born-when, died-when;
 - with other resources (like multimedia data);
 - **with mentioned manuscripts** (i.e. a person could be owner of a codex created by Vespasiano's school).

Example:

`<rdf:description rdf:about="LM">`

PSI: LM

Accepted form (Access key): Medici, Lorenzo de'

Variant forms: e.g. Lorenzo il Magnifico, Principe di Firenze, il Medici padre

Role: Principe

Biography:

1. http://it.wikipedia.org/wiki/Lorenzo_de'_Medici
2. in *Vite* (URL)

Relationships:

- Father-of: Piero de' Medici (people.rdf#PM)
- Born-in: Firenze

- Died-in: Firenze
- Born-when: 1449
- Died-when: 1492
- Owner-of: Vite di Plutarco (manuscripts.xml#PV1)

We can note that we have different kind of properties: some extracted from the letter itself (i.e. the variant forms of the name); some other that we can find via URL/URI (i.e. the biography); some that we have to deduce from other repertories (i.e. the name in a controlled form); and finally some other that we deduce from reading the text (i.e. different kind of relationships).

Starting from TEI P5 section on "Prosopographical Data" [6] it's possible to define a first analytic description of the features regarding persons.

Using also the VIAF³ for the access point to the control of authority, it's easy to define connection with standard systems and relation with already existent authority records: LCA⁴ and MARC expression for features of bibliographic data⁵.

A starting possible open prosopography [see also 1] for XV century people (today not existent) could be defined too – reflecting also on the conceptual model of the *The Prosopography of Anglo-Saxon England*⁶ -, eventually public and "populable" by researchers.

One aspect of interest is to create an open authority list and a collection of authority records in which researchers interested in defining profiles of people lived during the XV century could put

³ *Virtual International Authority File* <http://viaf.org>.

⁴ *Library of Congress Authorities* <http://authorities.loc.gov>.

⁵ *Machine-Readable Cataloging* <http://www.loc.gov/marc>.

⁶ <http://www.pase.ac.uk>.

other information about the existent profiles, but also to create new ones. Giving to researchers, in a crowdsourcing perspective, an existent structure of possible features, useful in order to describe one person, the collection of profiles could be realized in a shared environment. A sort of open, collaborative community for creating a prosopography of XV century people and an authority list for defining the access points to the personal name.

b) Manuscripts = manuscripts.rdf <rdf:description rdf:about="value-of-related-URI">

With regard to the texts mentioned in the letters it is possible to create link to other repertories and establish relationships with both persons and terms. So it is necessary define some properties:

- normalize the entry (author, title) – using a controlled vocabulary (e.g., the *Gnomon thesaurus*⁷);
- specify the shelfmark if the text is identified in an existent codex;
- relationships:
 - with the codex repository
 - with the codicological description
 - with a digital image
 - with the digitalized full text
 - **with a person relevant to it** (i.e., the owner, the requester, the copyist, etc.)

⁷ <http://www.gnomon.ku-eichstaett.de/Gnomon/en/Gnomon.html>.

Example:

`<rdf:description rdf:about="PV_1">`

PSI: PV_1

Accepted form (Access key): Plutarchus, *Vitae Parallelae*

Shelfmark: Laur. Plut. 65, 26

Relationships:

Available-in: Biblioteca Medicea Laurenziana, Firenze

- Described in: <http://opac.bml.firenze.sbn.it/Record.htm?idlist=1&record=710312453859>
- Item-in(image):
<http://teca.bmlonline.it/TecaViewer/index.jsp?RisIdr=TECA0000777138>
- Item-in (full-text): --
- Owned-by: people.rdf#LM

In this case we note, as for the "person" category, that we have to use information we can find inside the text of the letters (the codex manuscript), and we have to extract to be presented in a formal way (identifying an accepted form); for some other information we have to check on the WWW; some other more information are derived from the research knowledge (i.e. the shelfmark). The same situation we can notice for the lexicon that follows.

c) Lexicon = lexicon.rdf <rdf:description rdf:about="value-of-related-URI">

The analysis of the technical lexicon used by Vespasiano is an interesting exploration in the history of the book and in the actual trade of the copyists, and it is sometimes connected to the manuscripts he realized. At this level it is possible to give:

- definitions:
 - new ones;
 - present in other repertories, with link to the external entries;
- **relationships:**
 - with terms, that is define an internal thesaurus (i.e. "in foglia d'oro" is a type-of illumination, "cordovani" is a type-of binding);
 - with existent lexical ontologies;
 - **between manuscripts and lexicon** used.

Example:

<rdf:description rdf:about="aF">

PSI: aF

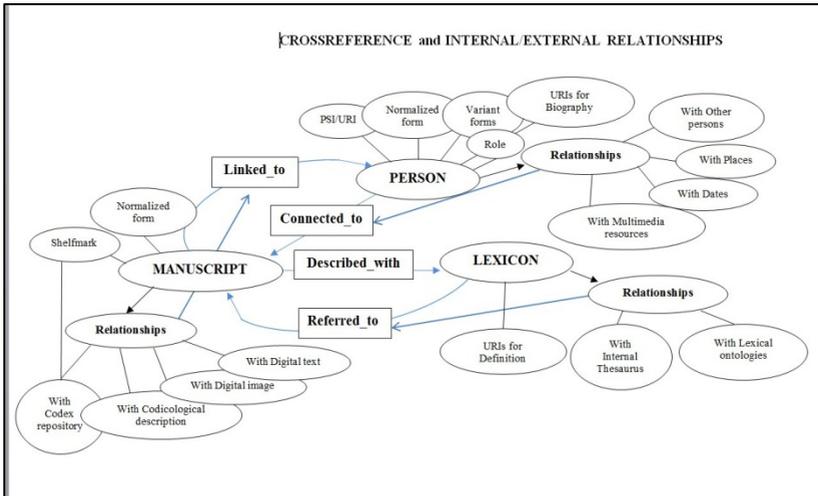
Accepted form (Access key): Ad fragmenta

Definition: "Sistema di resa scrittorea opposto alla scrittura 'ad volumina'. [...] (S. Rizzo, *Lessico filologico degli umanisti*, Roma, Edizioni di Storia e Letteratura, 1973, p. 196)".

Relationships:

- Type-of: writing
- Iperonym: scrittura
- Synomin: pecia
- Antonym: ad volumina
- **Referred-to: manuscripts.rdf#PV_1**

Figura 2: *Crossreference and internal/external relationships*



The knowledge base

All the external information was described in a formal way through the RDF model (see also *Henry the III Fine Rolls Project* [2, 7]⁸). Each property of each of the three categories (persons, manuscripts, lexicon) was described with appropriate values. All of the properties were represented as relationships, considering the three categories like classes of our ontology (i.e. the class person has the properties: PSI, accepted form, role, biography, etc. and specific values). But mostly we have to reflect on the different kind of relationships that each class has with the other two. We need to focus on the fact that between the instances of these classes we can define relationships that provide greater conceptual depth and that

⁸ <http://www.finerollshenry3.org.uk/cocoon/frh3/index.html>.

can be easily expressed in a formal language, creating a good model for the representation of the content of the letters. We can establish unambiguously the different relationships existing between a person, a manuscript and a term, such as a person being the owner, the copyist, or the client of a codex and this manuscript is described with a specific term (p.e. Piero de' Medici is owner of Plutarco's Vite, which are realized in "chordovani"). Integrating the specific created class with predicates defined in existent models, that we are studying, a suitable ontology could be distributed.

We are now analyzing different predicates for internal connection, starting from the principal existent ontologies. A first step in mapping our classes with the same classes in other ontologies – principally FRBR⁹, DC terms¹⁰, SKOS¹¹ and CIDOC-CRM¹² – was realized. The same operation was made on the properties. We consider also the TEI Ontologies SIG¹³ for studying other mapping models, starting from TEI [4].

Here some considerations based on the "is-related-to" property:

- Class PERSON (the Person/Agent).
- Relationships/properties: owner-of; copyist-of; illuminator-of; requester-of a manuscript.
- Class MANUSCRIPT (the Bibl/Object/Event). Relationships /

⁹ *Functional Requirements for Bibliographic Records*. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

¹⁰ *Dublin Core Terms* <http://dublincore.org/documents/dcmi-terms>.

¹¹ *SKOS Simple Knowledge Organization System Reference* <http://www.w3.org/TR/skos-reference>.

¹² *Cidoc- Conceptual Reference Model* <http://www.cidoc-crm.org>.

¹³ <http://www.tei-c.org/SIG/Ontologies>.

properties: created-for; requested-by; copied-by; illuminated-by a person / described-with terms.

- LEXICON (the Term/Concept). Relationships/properties: referred-to a manuscript.

But, like we explain in the previous section, the external information could be linked to some external resources in order to adding depth to the descriptions and to create a connection with Web resources. We are studying some predicates:

- PERSON. TEI prosopography integrated with CIDOC-CRM for relationships with events, dates and places.
- MANUSCRIPT. An ontology properly extending and customizing the *Functional Requirements for Bibliographic Records* (FRBR) can be used to capture the subtle difference between the physical codex, the content it has (which could be shared with other publications including electronic copies of it) and the work (e.g., the "Vite di Plutarco") the codex is a manifestation-of.
- LEXICON. The *Simple Knowledge Organization System* (SKOS) provides a basic ontological foundation for a terminological thesaurus and the relationships of the terms defined therein and elsewhere. As such, it can be used in a simple way to provide support for the lexicon data set.

But we mostly need to reflect on the fact that the texts mentioned are specific codices that can be found nowadays in a specific library, or a person has an iconographic representation could be found in a certain cultural institute. With URI reference and RDF representation we started to create relationships between people, manuscripts and lexicon of Vespasiano's letters and the related

concept all over the Web. For this reason we are now studying Linked Data¹⁴ for manage the external link [5]. Some linked datasets already exposed could be used¹⁵. Managing our information thought RDF and URI we can create a dataset that could be used by other projects, using already existent metadata and vocabularies for interchange and interoperability¹⁶.

Conclusions

The aim of the creation of an external knowledge is double: on the one hand it guarantees interoperability of the edition (it still is open to any dialogue with other repositories and systems) using unmodified standard languages such as XML/TEI (and the Dublin Core in order to expose metadata in a Web interface); on the other hand it gives scholars a useful tool (authority lists expressed in a formal way and accessed via URIs) and a method (from the content of the letters to external information and from the external information to the text) that can be used and reused in a number of different contexts and situations. Three levels of the knowledge base was defined: internal relationships, relationships between the internal annotation and other Web resources, internal linked data set and dialogue with other linked data sets. A model suitable and easily exportable in other projects, using RDF, URI and existent ontologies.

¹⁴ <http://linkeddata.org>.

¹⁵ SWEO Community Project: *Linking Open Data on the Semantic Web* <http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets>.

¹⁶ Library Linked Data Incubator Group: *Datasets, Value Vocabularies, and Metadata Element Sets* <http://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025>.

Riferimenti Bibliografici

- J. Bradley, S. Harold, "Texts into databases: the evolving field of new-style prosopography". In *Literary and Linguistic Computing*, 2005(20), suppl. 1, p. 3-24. doi:10.1093/lc/fqi022.
- Ciula, P. e V. Spence, "Expressing Complex Associations in Medieval Historical Documents: The Henry III Fine Rolls Project". In *Literary and Linguistic Computing*, 23, 2008(3), p. 311-325. doi:10.1093/lc/fqn018.
- J. Cummings, "The William Godwin's Diaries Project". In *Jahrbuch für Computerphilologie*, 2008(10), <http://computerphilologie.de/jg08/cummings.pdf>.
- Ø. Eide, Ch. E. Ore, *TEI, CIDOC-CRM and a possible interface between the two*, in *Digital Humanities 2006. The First ADHO International Conference: Conference Abstracts*, Paris, Université Paris-Sorbonne, 2006, <http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>.
- T. Heath, Ch. Bizer, *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011. doi: 10.2200/S00334ED1V01Y201102WBE001.
- TEI Consortium (ed.), *13.3 Biographical and Prosopographical Data*, in *Guidelines for Electronic Text Encoding and Interchange*, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDPERS>. (Last updated on 21 December 2011)
- J. M. Vieira, A. Ciula, *Implementing an RDF/OWL Ontology on Henry the III Fine Rolls*, Paper presented at OWLED 2007, Innsbruck, June 6-7, http://www.webont.org/owled/2007/PapersPDF/submission_6.pdf.

**Oltre l'informatica umanistica:
culture digitali e reti della conoscenza**

Per un critica culturale delle Digital Humanities

Domenico Fiormonte

Forme della crisi

L'informatica umanistica (IU) italiana si è caratterizzata fortemente, fin dall'inizio, per il suo contributo teorico. Grazie soprattutto a studiosi come Tito Orlandi, Raul Mordenti, Dino Buzzetti e naturalmente Giuseppe Gigliozzi, al centro del dibattito teorico-metodologico è stata posta la questione epistemologica che possiamo sinteticamente riassumere con la domanda: che tipo di conoscenza costruiamo con gli strumenti digitali e *come* la costruiamo? Questa domanda è all'origine del solido contributo che l'IU italiana (e romana) ha dato nel settore degli studi sulla codifica. La versione per così dire ermeneutica della IU da noi ha solo vagamente attecchito, essendosi poco diffusa l'analisi testuale con strumenti informatici. Per una serie di fattori, probabilmente inevitabili o comunque difficilmente contrastabili, il rigore di questa discussione ha avuto uno scarso impatto al di fuori degli addetti ai lavori, soprattutto a livello delle iniziative umanistiche (e i risultati a mio parere si vedono: centinaia di progetti di digitalizzazione di

biblioteche, archivi, ecc. operate con dubbio criterio pratico e scarsa consapevolezza teorica). Oggi constatiamo una crisi della spinta propulsiva di questi studi, a fronte però di un'estesa colonizzazione materiale e simbolica da parte delle tecnologie digitali. Digitalizzare non è più solo una moda o un imperativo: è diventato la normalità. Ma analizzare le cause esterne di questa situazione di stallo (lo ha fatto già Orlandi in varie sedi¹) è utile fino a un certo punto. A mio parere se vogliamo rilanciare questi studi e ridare fiato allo sforzo di ricerca fatto in questi anni dobbiamo partire anche dalle sue lacune, che non sono solo italiane. In questo intervento cercherò di metterne in evidenza alcune, senza proporre né ricette né soluzioni universali.

La prima lacuna consiste nella scarsa propensione e in taluni casi incapacità da parte delle DH di sviluppare quello che Pierre Bourdieu chiama «un modello teorico [che consiste nel] sottoporre a critica riflessiva gli strumenti con i quali si pensa la realtà»². Ovvero si costruiscono strumenti, si riflette sul loro uso e a volte sul loro impatto, ma ciò che è a monte, il fondamento culturale di questi strumenti, viene solo raramente considerato. In altri termini le DH partirebbero sempre dai "risultati" – senza interessarsi al processo che ci ha condotto sin lì. Un recente intervento di Alan Liu esprime

¹ Per esempio: <http://rmcisadu.let.uniroma1.it/~orlandi/pubbli/informatica/montevarchi.pdf>.

² P. Bourdieu, R. Chartier, *Le sociologue et l'historien*, Paris, Editions Agone et Raison D'Agir, 2010 (trad. it. *Il sociologo e lo storico. Dialogo sull'uomo e la società*, Bari, Dedalo, 2011), p. 47.

preoccupazioni molto simili sul «lack of cultural criticism»³ delle DH, invocando, con un'immagine molto appropriata (e in aperto contrasto con le tendenze *mainstream*), la fondazione di una «intellectual infrastructure» per le *digital humanities*. L'articolo di Liu, che sottolinea fra l'altro i limiti 'politici' dell'approccio strumentalista, è una sana boccata di ossigeno nel contesto anglo-americano, ma tuttavia non entra nel dettaglio degli equilibri geopolitici e degli interessi economici che agiscono all'interno del sistema delle DH.

La scarsa propensione a riflettere sull'origine dei propri oggetti ha probabilmente radici diverse, anche se non v'è dubbio che il carattere storico delle discipline umanistiche ha comportato un eccesso di concentrazione sulle procedure di conservazione, gestione e analisi dei dati, trascurando la componente forse più rivoluzionaria (in senso sia positivo sia negativo) dell'informatica, cioè la capacità di incidere sui *processi*, ancora prima che sui *prodotti* della ricerca.

L'altro limite, di carattere più concreto e a valle di quanto appena detto, è quello della composizione geopolitica e linguistico-culturale della disciplina e di conseguenza degli strumenti utilizzati⁴. Qui i problemi, sebbene intrecciati, sono di due ordini: a) composizione degli organi di governo delle istituzioni/ consorzi ecc.

³ A. Liu, "Where Is Cultural Criticism in the Digital Humanities?". In: M. K. Gold (ed.) *Debates in The Digital Humanities*, Minneapolis-London, University of Minnesota Press, 2012, p. 492.

⁴ Cfr. D. Fiormonte, "Il dibattito internazionale sull'informatica umanistica: formazione, tecnologia e primato delle lingua". In *Testo e Senso*, 4-5, (2001-2002), p. 145-156, <http://testoesensoold.uniroma2.it/article/show/34/il-dibattito-internazionale-sullinformatica-umanistica-formazione-tecnologia-e-primato-delle-lingue>.

che ispirano, gestiscono e regolano i processi, le strategie e le metodologie di ricerca (e la conseguente visibilità dei risultati); b) riflessi e caratteristiche linguistico-culturali degli strumenti⁵. All'interno di questo secondo ordine possiamo identificare: b1) il problema intrinseco – di ordine soprattutto politico e economico – del software prodotto o *copyrighted* quasi esclusivamente nel mondo anglo-americano; b2) il problema cosiddetto semiotico-culturale dei diversi strumenti di rappresentazione: dalle icone delle interfacce grafiche agli standard come Unicode, dalla prossemica di *Second Life* ai concetti universalistici di usabilità, ecc.

Cercherò di dire qualcosa soprattutto su a) e b2), sacrificando per ragioni di spazio la questione, pur importante, del software b1). Ma procediamo con ordine, partendo dagli assetti istituzionali e organizzativi.

⁵ Cfr. D. Fiormonte, "Il testo digitale: traduzione, codifica, modelli culturali", in *Italianisti in Spagna, ispanisti in Italia: la traduzione. Atti del Convegno Internazionale* (Roma, 30-31 ottobre 2007), a cura di P. R. Piras, A. Alessandro, D. Fiormonte, Roma, Edizioni Q, 2008, p. 271-284; Id., "Chi l'ha visto? Testo digitale, semiotica, rappresentazione in margine a un trittico di Dino Buzzetti". In *Informatica Umanistica*, 2009(2), p. 21-46.

Geopolitica della rete

Uno sguardo appena approfondito rivela che i condizionamenti del codice sono in generale altamente pervasivi, proprio perché introiettati come norma indiscussa. Un esempio banale è la lunga dittatura del codice ASCII a 7 bit (American Standard Code for Information Interchange), che è stato per oltre quaranta anni l'unico set di caratteri riconosciuto da tutte le piattaforme, ivi incluso il Web. A partire dal 2006 l'Internet Engineering Task Force (IETF) e l'Internet Corporation For Assigned Names and Numbers (ICANN) hanno lanciato il progetto Internationalized Domain Name⁶, la cui applicazione però, almeno nei paesi che utilizzano l'alfabeto latino, è ancora limitata (lo dimostra la ancora netta prevalenza del vecchio stile non-accentato nella scrittura delle URL). Il ritardo con cui queste basilari procedure di internazionalizzazione sono state portate avanti non dipende ovviamente da limiti tecnici, ma da scelte politiche: fino al 2009 l'ICANN, che si definisce quasi ossimoricamente una «not-for-profit public-benefit corporation», è stata sotto il controllo del Department of Commerce degli Stati Uniti⁷ e l'attuale CEO e presidente è Rod Beckstrom, già direttore del National Cybersecurity Center (NCSC) del Department of Homeland Security⁸. Un pedigree

⁶ <http://www.icann.org/en/news/announcements/announcement-30oct09-en.htm>.

⁷ http://www.readwriteweb.com/archives/commerce_department_loosens_grip_on_icann.php.

⁸ <http://www.icann.org/en/biog/beckstrom.htm>.

impeccabile per un cyberpoliziotto, un pò meno per un manager di un bene comune come la Rete⁹.

Venendo a settori più vicini agli interessi umanistici, gli assetti di potere non appaiono meno sconsolanti. Un esempio è il caso Unicode (un cui caso specifico analizzerò fra poco). Unicode, sulla carta, è una organizzazione non-profit «devoted to developing, maintaining, and promoting software internationalization standards and data, particularly the Unicode Standard, which specifies the representation of text in all modern software products and standards»¹⁰. Il Board of Directors è attualmente composto da due dipendenti di Google, due di Microsoft, uno di Apple, uno di JustSystems, uno di IBM e uno di OCLC¹¹. Non dissimile la composizione degli Executive Officers: il presidente è un ingegnere di Google dal 2006 e, a parte un paio di eccezioni provenienti dal mondo accademico o della ricerca, nessuna istituzione pubblica è rappresentata. Di fatto Unicode è uno standard industriale controllato dall'industria. Qualsiasi discorso (o lamento) sulla 'terzietà' o effettiva indipendenza di questi organismi è destinato ad arenarsi su questi dati

⁹E tuttavia rispetto a questa concentrazione in mani occidentali delle principali organizzazioni di gestione e controllo della rete, i dati dell'accesso a Internet mostrano un quadro ben diverso: il mondo cosiddetto occidentale (Europa e USA) rappresenta solo il 35.7% del totale degli utenti, mentre l'Asia il 44%. (Fonte: <http://www.internetworldstats.com/stats.htm>).

¹⁰ <http://www.unicode.org/consortium/consort.html>.

¹¹ <http://www.unicode.org/consortium/directors.html>.

Se tali sono gli scenari globali, è chiaro che la situazione delle DH non può non risentirne. Rispetto all'analisi da me condotta nel 2002 sono stati fatti grossi sforzi per rendere più internazionali e più multiculturali le varie associazioni e organizzazioni, ma l'impressione è sempre di una solida base anglo-americana sulla quale si innestano qui e lì individualità di altri paesi (per altro europei). Iniziative come Centernet, ADHO (Alliance of Digital Humanities Organizations¹²) e per ultima CHAIN (Coalition of Humanities and Arts Infrastructures and Networks) hanno il merito di raccogliere e censire le maggiori realtà dell'asse atlantico USA-Canada-UK, ma si tratta di operazioni di auto-rafforzamento delle identità già esistenti piuttosto che di reale condivisione delle conoscenze o effettiva esplorazione di culture, metodologie e pratiche altre. Lo testimonia non solo lo sconsolante monolinguisimo di questi siti, dove la struttura retorica non lascia spazio ad altro immaginario che non sia quello 'internò anglo-americano¹³, ma l'auto-denuncia di certe iniziative ascientifiche, e tuttavia eloquenti, come la *flatlandia* realizzata da Melissa Terras, un *infographics* in cui le digital humanities appaiono come un impero composto da due macroregni, USA e UK, e intorno qualche sparuto satellite (Fig. 1). Sono rappresentazioni (o auto-rappresentazioni) universalistiche come questa che mettono a nudo le condizioni di subalternità nelle

¹² Attualmente lo Steering Committee di ADHO è composto da sette membri, di cui solo due non provenienti da USA o UK: <http://digitalhumanities.org/administration/steering>.

¹³ Cfr. A. S. Canagarajah, *A Geopolitics of Academic Writing*, Pittsburgh, University of Pittsburgh Press, 2002.

soprattutto nelle scienze umanistiche e sociali, delle *categorie residuali*:

Qualsiasi tentativo di creare un sistema di classificazione obbligatorio, rigido ed universale darà come risultato, immediatamente, categorie residuali (...) bisogna radicare la consapevolezza di ciò che accade ogni qual volta si tenta di standardizzare, ovvero che in questa creazione c'è chi vince e c'è chi perde. Non si tratta di una questione semplice, né tanto meno di una materia facile da analizzare¹⁶.

Il problema della crisi (che è anche di autostima) delle scienze umanistiche¹⁷ potrebbe essere riassunto tutto qui: la necessità costitutiva, per continuare a esistere, di essere sempre e comunque margine, ibrido o variante del sistema. Ed è qui che sorge il primo ostacolo, il potenziale attrito fra il ruolo delle 'DH' e quello delle 'H' (senza la D); perché è evidente che un rilancio, o almeno una rivitalizzazione che non sia semplice difesa a oltranza dell'esistente, non può non realizzarsi senza una critica degli interessi economici e geopolitici che si celano dietro l'universo della rete e delle sue applicazioni. Le DH sono dunque vittima di un continuo paradosso: dimostrare di essere all'altezza della 'tecnica' (e dei suoi protagonisti, reali e virtuali) e allo stesso tempo evitare di diventarne succubi.

¹⁶ G. Bowker, S. Leigh Star *Sorting Things Out. Classification and its Consequences*, Cambridge-London, The MIT Press, 1999, p. 13.

¹⁷ Il legame fra crisi delle scienze umanistiche e ruolo delle DH è il tema dell'iniziativa di *advocacy* portata avanti da un gruppo di università, associazioni e centri di ricerca britannici, americani, canadesi e australiani: <http://humanistica.ualberta.ca>.

Standard ed egemonie culturali

Secondo G. Bowker e S. Leigh Star «classifications and standards are material, as well as symbolic», e il loro controllo «is a central, often underanalyzed feature of economic life»¹⁸. Nel loro studio i due sociologi mostrano come le tecniche di classificazione (e gli standard collegati o da essi generati) abbiano sempre svolto un rilevante ruolo di indirizzo economico e socio-culturale. Gli attuali standard delle tecnologie digitali sono frutto di un doppio *bias*: quello tecnico e quello culturale (geopolitico). I due *bias* sono intrecciati e non è quasi mai possibile distinguere dove inizia la scelta tecnologica e finisce il pregiudizio culturale.

Come notava il lessicografo e blogger José Antonio Millán più di dieci anni fa «se le reti sono le autostrade dei flussi di beni e servizi digitali, le tecnologie legate alle lingue degli utenti sono i pedaggi obbligatori»¹⁹. Dunque alle radici della supremazia economica, politica e culturale non vi è la tecnologia "bruta", ma piuttosto il controllo che questa garantisce sulle lingue: una serie di algoritmi, protocolli e software quasi sempre protetti dalle ferree leggi del copyright che ne gestiscono e ne controllano lo sviluppo. Gli standard allora non sono altro che il risultato di rapporti di forza²⁰. Presidiare le tecnologie linguistiche è insieme un succulento

¹⁸ Ivi, p. 15 e 39.

¹⁹ J. A. Millán, *Internet y el español*, Madrid, Retevision, 2001, p. 140.

²⁰ «La nostra nuova infrastruttura informativa globale, d'altro canto, è basata su schemi di classificazione elaborati nei paesi sviluppati per venire incontro ai problemi soprattutto dell'élite istruita.» (G. Bowker, S. Leigh Star, "Intervista su società dell'informazione e disuguaglianze", In. *Daedalus*, 2006(19), p. 15, <http://epl.scu.edu/~gbowker/interview.pdf>).

business e una mossa geopolitica. Come ricorda Millán, per molti paesi, Italia inclusa, non investire in questo settore vorrà dire, in un futuro ormai alle porte, essere forzati a pagare per poter utilizzare la propria lingua.

Ma il problema dell'egemonia culturale non è certo solo una questione linguistica. L'universo tecnologico-informatico, almeno da quaranta anni, è dominato da rappresentazioni culturali e metafore importate dall'universo anglo-americano. Molti elementi divenuti ormai familiari provengono dall'immaginario quotidiano USA e sono stati assorbiti dall'industria informatica. Non stiamo parlando di linguaggi di programmazione e algoritmi, dove il profondo *bias* semiotico è evidente²¹, ma del rutilante e non meno sottile universo delle icone divenuto lo standard delle GUI (interfacce grafiche). Per esempio la familiare cartellina gialla con la linguetta, tipica degli uffici americani, che grazie prima allo Xerox Star desktop²² e poi a Mac diventerà la metafora del «contenuto» dei singoli documenti del nostro computer. Si tratta del *manila folder*, dove «manila» si riferisce ai manufatti di fibra canapa ricavata dalle palme delle Filippine. Un residuo coloniale divenuto ubiquo, un simbolo che marca (e cela) fino a oggi l'origine "burocratica" del computer e il suo legame con l'immaginario USA. Un altro esempio della colonizzazione cybersimbolica sono le espressioni facciali e le animazioni (*playable*

²¹ Cfr. P. B. Andersen, *A Theory of Computer Semiotics. Semiotic Approaches to Construction and Assessment of Computer Systems*, Cambridge, Cambridge University Press, 1997.

²² «By far its most striking feature was its graphical user interface, ... The arrangement of folders and icons built around what the Star engineers called the 'desktop metaphor' is so familiar today that it seems to have been part of computing forever.» (HILZIK: 1999, 364).

animations) di *Second Life*, dove gli avatar hanno a disposizione una 'bachecà di espressioni e gesti che potrebbero essere decifrati solo da esperti parlanti nativi.²³ Si prenda per esempio il famoso «kiss my butt» (Figura 2), dove sia l'espressione verbale, sia la postura, in molte culture latine o mediterranee, potrebbero suggerire significati non solo totalmente ingannevoli ma addirittura opposti a quello originario (pressappoco «vai a quel paese»).

Figura 2. Il gesto del «kiss my butt» in *Second Life*



²³ Una lista complete di queste animazioni si trova su: http://wiki.secondlife.com/wiki/Internal_Animations#User-playable_animations.

La lista degli insediamenti iconici americani potrebbe continuare all'infinito, ma vediamo ora un esempio di *bias* rappresentazionale che riguarda direttamente il lavoro degli umanisti. Mi riferisco all'imponente lavoro portato a termine dal consorzio Unicode sulla codifica dei caratteri.

Antonio Perri ha fornito alcuni esempi convincenti del bias delle tecnologie di rappresentazione dei caratteri, mostrando come esistano rischi concreti di un «inaridimento della ricchezza fenomenologica delle prassi scritte umane»²⁴. Le soluzioni di codifica proposte dal Consorzio Unicode per problematiche diverse riguardanti le scritture del subcontinente indiano, il cinese, l'arabo e l'*hangul* (la scrittura coreana), oltre a essere eccessivamente dipendenti dai software di visualizzazione (mettendo a repentaglio proprio la trasmissibilità), rivelano un condizionamento fondato su una visione «iper-tipografica» (cioè alfabeto-centrica) della scrittura che trascura le funzionalità degli aspetti «visiografici»²⁵. Perri fornisce alcuni convincenti esempi di tale *bias* quando discute la codifica Unicode delle legature e della posizione delle vocali nella scrittura indiana Devanagari. Nei sistemi indiani «spesso fattori di ordine grafico hanno il sopravvento sull'ordine di lettura dei grafemi». Come si vede nella Figura 3, nel secondo glifo (p+i) l'ordine pronuncia / sequenza grafica è invertito. Gli esperti di Unicode tuttavia sostengono che Unicode

²⁴ A. Perri, "Al di là della tecnologia, la scrittura. Il caso Unicode". In *Annali dell'Università degli Studi Suor Orsola Benincasa*, 2009(II), p. 747.

²⁵ Ivi, p. 729.

[...] codifica le scritture indiane secondo uno schema logico, ignorando i dettagli tipografici come le legature o l'esatto posizionamento dei caratteri vocalici...». Ma perché mai l'ordine dei caratteri corrispondente al segmento fonico dovrebbe essere logico per uno scrivente indiano? Chi ci dice che la linearità del significante saussuriano-alfabetico debba giocare un ruolo nelle *sue* pratiche di scrittura? E perché mai le legature e la struttura visivo-compositiva del testo scritto indiano dovrebbero essere considerate meri *dettagli tipografici*?²⁶

Figura 3. Due grafemi Devanagari (Perri 2009, 735).

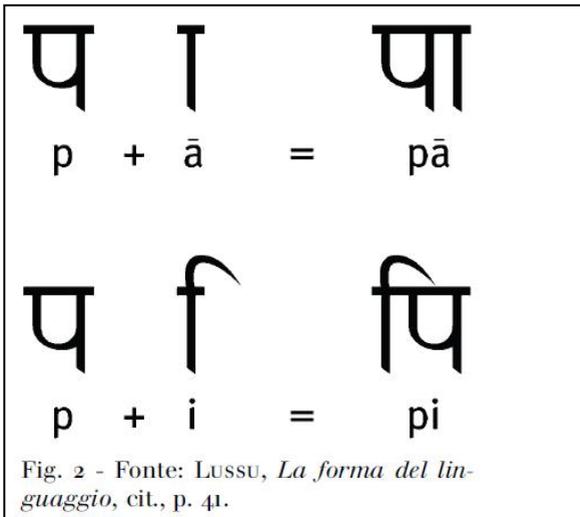


Fig. 2 - Fonte: LUSSU, *La forma del linguaggio*, cit., p. 41.

²⁶ Ivi, p. 736.

Diversità linguistica e ineguaglianze globali

L'ultimo esempio in realtà non un vero e proprio esempio, ma piuttosto un esperimento comparativo basato su due rappresentazioni grafiche. La prima immagine (Figura 4) è una mappa delle diseguaglianze del mondo tratta dal database dello University of California Atlas of Global Inequality. La seconda mappa (Figura 5), è un'immagine tratta da Wikipedia elaborata sulle fonti disponibili su Ethnologue.com e rappresenta la diversità linguistica nel mondo. In rosso sono indicate le otto mega-regioni che insieme raggruppano più del 50% delle lingue mondiali, in blu sono indicate le aree di grande diversità.

Figura 4. Prodotto interno lordo (GDP= Gross Domestic Product) mondiale nel 2004. Fonte: <http://ucatlas.ucsc.edu/>

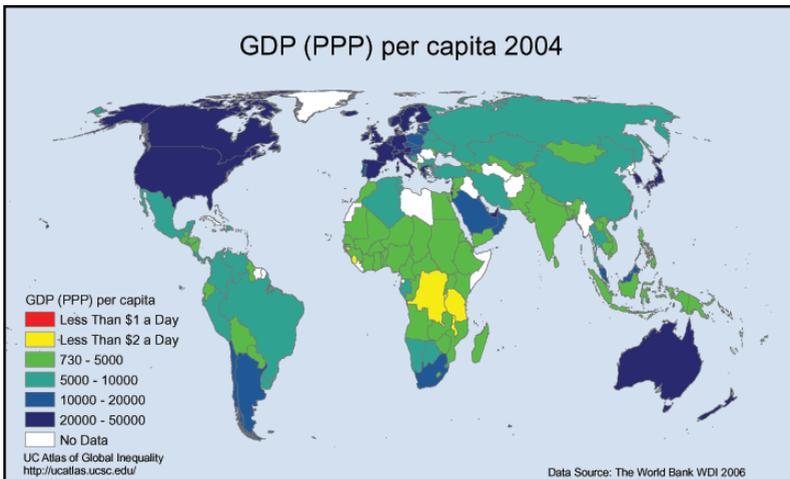
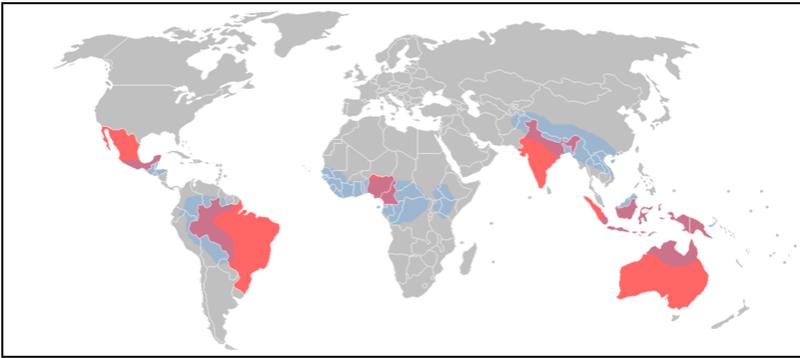


Figura 5. Diversità linguistica nel mondo²⁷



Se queste due mappe vengono sovrapposte, escluse alcune zone come l'Australia dove la forte diversità linguistica è dovuta soprattutto all'alto numero di immigrati dal resto del mondo²⁸, si può notare come le zone a più basso reddito coincidano in molti casi con le aree dove è maggiore la diversità (e dunque la ricchezza) linguistica. L'accostamento fra le due prime mappe non vuole e non può suggerire facili conclusioni, tuttavia è lecito ipotizzare che in molte fra le zone più povere del mondo, ovvero nei deserti, nelle giungle e nelle montagne ai margini della globalizzazione, resistano popolazioni che nella miseria forse coltivano l'unica possibile ricchezza di cui dispone ancora l'uomo: la parola.

²⁷ Fonte: http://en.wikipedia.org/wiki/Linguistic_diversity#Linguistic_diversity.

²⁸ L'invasione europea dei secoli XIX e XX sradicò le culture e le lingue delle popolazioni aborigene, tuttavia secondo l' Australian Bureau of Statistics "Today, there are approximately 22 million Australians, speaking almost 400 languages, including Indigenous languages". (<http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1301.0Feature%20Article32009%E2%80%93310?opendocument&tabname=Summary&prodno=1301.0&issue=2009%9610&num=&view>)

Non sorprende, infine, che la mappa delle diseguaglianze sia sovrapponibile anche a quella della diffusione delle DH mostrata sopra (cfr. Fig. 1): il che non solo conferma l'ipotesi di Millán sul rapporto fra dominio economico, concentrazione tecnologica e uniformazione linguistica, ma solleva la questione sinora mai affrontata di un *digital humanities divide* interno ed esterno ai paesi occidentali.

Conclusioni

Qual è il ruolo e la posizione delle DH rispetto agli scenari geopolitici che abbiamo abbozzato? La mia impressione è che, nonostante l'indubbia espansione di questa disciplina²⁹, le DH non abbiano ancora preso atto di quanto sia avvenuto e stia avvenendo a livello globale. Ma forse si tratta dell'inevitabile rimozione di una verità troppo amara per essere contemplata: le *digital humanities* in versione globalizzata, fino ad oggi, non hanno né rafforzato le scienze umanistiche né riequilibrato i rapporti di forza fra queste e le scienze "dure".

La prospettiva dei «formal methods»³⁰, se da un lato non è riuscita a instaurare un dialogo paritario con l'informatica, dall'altro

²⁹ Cfr. M. K. Gold, "The Digital Humanities Moment". In M. K. Gold (ed.) *Debates in The Digital Humanities*, , Minneapolis-London, University of Minnesota Press, 2012, p. IX-XVI.

³⁰ Per una aggiornata discussione sul tema della formalizzazione nelle discipline umanistiche, cfr. J. Van Zundert, S. Antonijevic, A. Beaulieu, K. Van Dalen-Oskam, D. Zeldenrust, T. L. Andrews, "Cultures of Formalisation: Towards an Encounter between Humanities and Computing", In D. M. Berry (ed.) *Understanding Digital Humanities*, , Houndmills; New York, Palgrave Macmillan, 2012.

ha finito per rendere l'informatica umanistica abbastanza miope, se non ostili, nei riguardi della cosiddetta 'cultura digitale', rilegandola a questione 'sociologica. Come già sottolineava negli anni Novanta la ricerca ACO*HUM³¹, e ha ripetuto implicitamente Buzzetti nel suo intervento, il computer è una "macchina universale" e l'applicazione di metodi formali è il minimo comun denominatore delle DH. Buzzetti ha posto il problema fra teoria e tecnologia, io qui ho tentato di porre quello fra tecnologia e cultura materiale. Per Buzzetti, ad esempio, l'aspetto grafico occulta l'aspetto logico, ma questo perché egli è un logico che privilegia questo stile cognitivo; tuttavia un testo o meglio un artefatto cognitivo-culturale non è solo il riflesso di una struttura logica, ma anche di un'interazione dinamica, di un significato che emerge dalla cooperazione e non solo dall'analisi o dalla modellazione. Insomma: un artefatto culturale non è una struttura logica, non è solo sintassi e semantica, è anche pragmatica. È per questo che i dati della cultura umana sono difficilmente formalizzabili.

D'altra parte la dipendenza della macchina dall'alfabeto va ben oltre la difficoltà o incapacità di staccarsi dal modello della stampa e sarebbe iscritta nel suo stesso DNA:

[L]a macchina digitale è in primo luogo una macchina alfabetica, poi è logica e formale. [...] Poniamoci dunque il problema di considerare che impatto ha tale macchina nella costruzione di conoscenza; infatti, la macchina non è neutrale, impone a chi la usa

³¹ Il sito è ancora attivo: <http://www.hd.uib.no>. Cfr. K. De Smedt, H. Gardiner, E. Ore, T. Orlandi, J. Souillot, W. Vaughan, (ed.), *Computing in Humanities Education: A European Perspective*, Bergen, University of Bergen (HIT centre), <http://www.hd.uib.no/AcoHum/book>.

una storia ed una logica, uno sguardo organizzatore dei fenomeni.... Inizia con ciò un percorso, una cultura, ben diverso da quelli inerenti alla scrittura geroglifica, ideogrammatica, dove tramite il disegno si propone un concetto, si evoca un'intera immagine, una situazione, un sentimento.

L'alfabeto invece discretizza, suddivide il continuo del linguaggio in atomi insignificanti, in bits, le lettere³².

In conclusione, l'implicito appiattimento sulle politiche tecnologiche, commerciali e industriali da un lato e l'origine tendenzialmente monoculturale delle rappresentazioni simboliche e logiche sono, a mio parere, fra i principali ostacoli per un'espansione non meramente strumentale delle DH. Sono anche d'accordo con Alan Liu (2012) che per estendere il loro raggio d'azione e "legittimarsi" come disciplina le DH hanno bisogno di confrontarsi e contaminarsi con le discipline limitrofe³³; oltre alla galassia delle scienze sociali (dagli Science and Technology Studies³⁴ alla Mediologia³⁵) citerei l'antropologia culturale, nella variante che si è

³² G. Longo, *Critica della ragion informatica in scienze della natura*, Lezione Galileiana (Pisa 25 ottobre 2006), in *Pianeta Galileo* (2007), http://www.consiglio.regione.toscana.it/news-ed-eventi/pianeta-galileo/atti/2006/03_longo.pdf. È significativo che a lanciare questo *l'accuse* sia proprio un informatico e matematico impegnato ultimamente nella ricerca biologica. Uno speculare supporto storico-teorico alle tesi di Longo può essere offerto dagli studi sulle origini numeriche del cuneiforme dell'orientalista Denise Schmandt-Besserat (*How writing came about*, Austin, University of Texas Press, 1996).

³³ Avevo indicato una lista di possibili intersezioni in T. Numerico, D. Fiorimonte, F. Tomasi, F. *L'umanista digitale*, Bologna, Il Mulino, 2010, p. 102-103.

³⁴ Per quanto riguarda l'Italia si veda il sito <http://www.stsitalia.org/?tag=science-and-technology-studies>.

³⁵ Calco dal francese médiologie (<http://www.mediologie.org/>), il termine si è diffuso anche in Italia (<http://www.mediologia.com/>). Secondo Régis Debray la mediologia

sempre occupata degli artefatti culturali (da André Leroi-Gourhan a Jack Goody, dall'etnografia di James Clifford e George E. Marcus alla *digital ethnography* di Michael Wesch³⁶). Ma forse la questione più urgente, soprattutto riguardo la IU nostrana, è smettere di considerare la questione metodologica e la questione culturale come entità separabili. Ovvero che, parafrasando Harold Innis, che le digital humanities non nascono in un *vacuum*:

Innis happily accepted as a starting point the inevitably ethnocentric bias of social science. (...) He recognized that scholarship was not produced in a historical and cultural vacuum but reflected the hopes, aspirations, and heresies of national cultures. American and British scholarship was based, he thought, on a conceit: it pretended to discover Universal Truth, to proclaim Universal Laws, and to describe a Universal Man. Upon inspection it appeared, however, that its Universal Man resembled a type found around Cambridge, Massachusetts, or Cambridge, England: its Universal Laws resembled those felt to be useful by Congress and Parliament; and its Universal Truth bore English and American accents. Imperial powers, so it seems, seek to create not only economic and political clients but intellectual clients as well³⁷. (Carey 1992, 149).

«tratta, in prima approssimazione, di analizzare le 'funzioni sociali superiori' (religione, ideologia, arte, politica) nei loro rapporti con i mezzi e gli ambienti di trasmissione e di trasporto» (R. Debray, "Che cos'è la mediologia?". In *Le Monde Diplomatique-II Manifesto*, settembre 1999, <http://www.mondediplomatique.it/LeMonde-archivio/Settembre-1999/9909lm02.01.html>).

³⁶ <http://mediatedcultures.net>.

³⁷ J. W. Carey, *Communication as Culture. Essays on Media and Society*, New York-London, Routledge, 1992.

Riferimenti Bibliografici

- P. B. Andersen, *A Theory of Computer Semiotics. Semiotic Approaches to Construction and Assessment of Computer Systems*, Cambridge, Cambridge University Press, 1997.
- P. Bourdieu, R. Chartier, *Le sociologue et l'historien*, Paris, Editions Agone et Raison D'Agir, 2010 (trad. it. *Il sociologo e lo storico. Dialogo sull'uomo e la società*, Bari, Dedalo, 2011).
- G. Bowker, S. Leigh Star *Sorting Things Out. Classification and its Consequences*, Cambridge-London, The MIT Press, 1999.
- G. Bowker, S. Leigh Star, *Intervista su società dell'informazione e disuguaglianze*, a cura di G. Pellegrino, «Daedalus», 19, 2006, p. 13-20, <http://epl.scu.edu/~gbowker/interview.pdf>.
- A. S. Canagarajah, *A Geopolitics of Academic Writing*, Pittsburgh, University of Pittsburgh Press, 2002.
- J. W. Carey, *Communication as Culture. Essays on Media and Society*, New York-London, Routledge, 1992.
- R. Debray, *Che cos'è la mediologia?*, in «Le Monde Diplomatique-II Manifesto», settembre 1999, <http://www.monde-diplomatique.it/LeMonde-archivio/Settembre-1999/9909lm02.01.html>.
- K. De Smedt, H. Gardiner, E. Ore, T. Orlandi, J. Souillot, W. Vaughan, (eds.), *Computing in Humanities Education: A European Perspective*, Bergen, University of Bergen (HIT centre), <http://www.hd.uib.no/AcoHum/book/>.
- D. Fiormonte, *Chi l'ha visto? Testo digitale, semiotica, rappresentazione*. In margine a un trittico di Dino Buzzetti, in «Informatica Umanistica», 2, 2009, p. 21-46.

D. Fiormonte, *Per una critica culturale delle Digital Humanities*

- D. Fiormonte, Il dibattito internazionale sull'informatica umanistica: formazione, tecnologia e primato delle lingua, in «Testo e Senso», 4-5, (2001-2002, p. 145-156, <http://testoesensoold.uniroma2.it/article/show/34/il-dibattito-internazionale-sullinformatica-umanistica-formazione-tecnologia-e-primato-delle-lingue>).
- D. Fiormonte, Il testo digitale: traduzione, codifica, modelli culturali, in *Italianisti in Spagna, ispanisti in Italia: la traduzione. Atti del Convegno Internazionale (Roma, 30-31 ottobre 2007)*, a cura di P. R. Piras, A. Alessandro, D. Fiormonte, Roma, Edizioni Q, 2008, p. 271-284.
- D. Fiormonte, V. Matiradonna, D. Schmidt, Digital Encoding as a Hermeneutic and Semiotic Act: The Case of Valerio Magrelli, in «Digital Humanities Quarterly», 4, 2010, 1, <http://digitalhumanities.org/dhq/vol/4/1/000082/000082.html>.
- D. Fiormonte, D. Schmidt, La rappresentazione digitale della varianza testuale, in *Canoni liquidi*, a cura di D. Fiormonte, Napoli, ScriptaWeb, 2011, p. 161-180.
- M. K. Gold, The Digital Humanities Moment, in *Debates in The Digital Humanities*, ed. M. K. Gold, Minneapolis-London, University of Minnesota Press, 2012, p. IX-XVI.
- M. A. Hilzik, *Dealers of Lightning*, New York, HarperCollins, 1999.
- A. Liu, Where Is Cultural Criticism in the Digital Humanities?, in *Debates in The Digital Humanities*, ed. M. K. Gold, Minneapolis-London, University of Minnesota Press, 2012, p. 490-509.
- G. Longo, Critica della ragion informatica in scienze della natura, Lezione Galileiana (Pisa 25 ottobre 2006), in *Pianeta Galileo (2007)*, http://www.consiglio.regione.toscana.it/news-ed-eventi/pianeta-galileo/atti/2006/03_longo.pdf.

- J. A. Millán, *Internet y el español*, Madrid, Retevision, 2001.
- T. Numerico, D. Fiorimonte, F. Tomasi, F. L'umanista digitale, Bologna, Il Mulino, 2010.
- A. Perri, Al di là della tecnologia, la scrittura. Il caso Unicode, in «Annali dell'Università degli Studi Suor Orsola Benincasa», II, 2009, p. 725-748.
- D. Schmandt-Besserat, *How writing came about*, Austin, University of Texas Press, 1996.
- D. Schmidt, The Inadequacy of Embedded Markup for Cultural Heritage Texts, in «Literary and Linguistic Computing», 25, 2010, 3, p. 337-356.
- J. Van Zundert, S. Antonijevic, A. Beaulieu, K. Van Dalen-Oskam, D. Zeldenrust, T. L. Andrews, *Cultures of Formalisation: Towards an Encounter between Humanities and Computing*, in *Understanding Digital Humanities*, ed. D. M. Berry, Houndmills; New York, Palgrave Macmillan, 2012.

Web semantico, linked data e studi letterari: verso una nuova convergenza

Fabio Ciotti

La storia finora: luci e ombre delle Digital Humanities

Uno dei libri più influenti nella storia del rapporto tra informatica e scienze umane è stato senza dubbio il notissimo *Hypertext* di George Landow¹. In quel volume lo studioso statunitense sosteneva che l'allora incipiente tecnologia dell'ipertesto avrebbe favorito una feconda convergenza tra la tradizione degli studi letterari (almeno per come questi si erano venuti evolvendo negli ultimi decenni del secolo scorso) e il dominio delle discipline computazionali e dei nuovi media digitali.

Dalla pubblicazione della prima edizione di quel libro sono passati ormai venti anni, e possiamo dire che l'era dell'ipertesto è ormai alle spalle: da una parte l'introduzione e l'evoluzione del Web ha per molti versi banalizzato l'ipertesto, rendendolo una forma/tecnologia di organizzazione delle informazioni di uso

¹ G. P. Landow, *Hypertext: the convergence of contemporary critical theory and technology*, Baltimore, Johns Hopkins University Press, 1992.

comune, al contempo attenuandone molte delle caratteristiche per così dire rivoluzionarie; dall'altra ha spostato l'attenzione sulla condivisione sociale del sapere, sulla creatività diffusa e sulla cooperazione tra gli individui mediata dalle tecnologie di comunicazione digitale. Naturalmente in questo contesto di profonda trasformazione delle modalità di creazione disseminazione e conservazione della conoscenza anche i saperi umanistici, i loro attori e i loro oggetti, hanno giocato un ruolo importante.

In questo modo l'Informatica umanistica, o *Humanities Computing* nella formulazione anglosassone, si è venuta progressivamente liberando della stigmatizzazione di disciplina di nicchia, riuscendo al contempo a ottenere una presenza rilevante nella didattica offerta dalla facoltà umanistiche (e questo anche in Italia, nonostante ritardi, ritrosie culturali e crisi dell'università in generale abbiano senza dubbio rappresentato e ancora rappresentino fattori di ostacolo); a conseguire importanti risultati sul piano della ricerca; a promuovere e consolidare infrastrutture e organizzazioni per la cooperazione scientifica a livello nazionale e internazionale che raccolgono e coordinano un numero ormai grandissimo di studiosi a livello planetario, organizzano convegni mastodontici e pubblicano monografie e periodici autorevoli².

² Basti ricordare l'annuale conferenza «Digital Humanities», cui attendono centinaia di ricercatori, la storica rivista «Literary and Linguistic Computing», cui si sono aggiunte più recentemente «Digital Humanities Quarterly», «Text Technology» e «Digital Studies / Le champ numérique», e i due ponderosi volumi miscelanei editi dalla Blackwell: S. Schreibman, R. G. Siemens, J. Unsworth, *A companion to digital humanities*,

La recente e rapida diffusione del termine *Digital Humanities*³ sancisce sul piano linguistico il successo di questo processo di consolidamento e generalizzazione, che ha attirato l'attenzione anche delle redazioni culturali della grande stampa, come testimonia la serie di articoli scritti dalla giornalista del *New York Times* Patricia Cohen, nel primo del quale si legge:

Members of a new generation of digitally savvy humanists argue it is time to stop looking for inspiration in the next political or philosophical "ism" and start exploring how technology is changing our understanding of the liberal arts. This latest frontier is about method, they say, using powerful technologies and vast stores of digitized materials that previous humanities scholars did not have⁴.

Una semplice rassegna dei risultati più notevoli conseguiti in questo campo richiederebbe di gran lunga troppo spazio per pensare di proporla in questa sede. Ma in questo panorama, se lo osserviamo senza l'ausilio del cannocchiale, quali sono le vette che si stagliano con maggiore rilievo? Quali sono i risultati più importanti che la ormai pluridecennale attività di ricerca nel dominio delle Digital Humanities ha prodotto? A nostro parere possiamo riassumerli nei seguenti punti:

Malden, Mass., Blackwell Pub., 2004 e R. G. Siemens, S. Schreibman, *A companion to digital literary studies*, Malden, MA, Blackwell Pub., 2007.

³ Termine per il quale è invero arduo trovare una traduzione soddisfacente in italiano. Preferiamo pertanto mantenere la formulazione inglese.

⁴ P. Cohen, "Digital Keys for Unlocking the Humanities' Riches", *New York Times*, novembre 16, 2010. http://www.nytimes.com/2010/11/17/arts/17digital.html?_r=2&emc=eta1&.

- 1) La consapevolezza teorica e metodologica, ampiamente condivisa dai centri di elaborazione e dai singoli studiosi più avanzati e propulsivi della comunità, che vede nel rapporto con le metodologie informatiche un elemento epistemologica-mente e teoricamente rilevante e non un semplice fattore strumentale.
- 2) Il concetto di *modellizzazione* come attività intellettuale che caratterizza l'attività di indagine sugli oggetti e i fenomeni culturali mediante il computer, mediando tra il livello della teoria e quello dell'osservazione. Il livello al quale si colloca dunque l'informatica nella ricerca umanistica è quello propriamente del metodo, un metodo che è relativamente *theory-independent* ma che richiede alla teoria la qualità del rigore formale, l'esplicitazione degli enti teorici e delle relazioni tra tali enti che essa presuppone e l'indicazione di procedure per collegare tali enti ai dati osservativi (in ultima analisi al materiale linguistico testuale e a quello documentale o fattuale contestuale).
- 3) La predisposizione di linguaggi e standard condivisi per la modellizzazione, rappresentazione e disseminazione di risorse digitali di qualità, attività legata alla forte cooperazione con la comunità scientifica archivistica e biblioteconomica.
- 4) Le ampie campagne di digitalizzazione di fonti primarie e secondarie in formato testuale e/o immagine facsimilare e la predisposizione di vasti *repositories* on-line che ormai mettono a disposizione in forma libera e gratuita e con livelli di affidabilità linguistica piuttosto elevati una parte importante della tradizione testuale occidentale.
- 5) Lo sviluppo di importanti framework e infrastrutture software per effettuare *information retrieval*, analisi testuale e pubblicazione on-line di tali risorse testuali, in genere disponibili liberamente come prodotti *open source* o *web service*.

A fronte di questi risultati, tutti di vasta portata ma collocati sul piano dei fondamenti teorici e metodologici e su quello delle infrastrutture generali per la ricerca, sta invece la oggettiva limitatezza dei risultati specifici (fatti salvi alcuni meritevoli controesempi) sul piano dei singoli campi disciplinari. Insomma il movimento delle Digital Humanities ha prodotto una notevole mole di risorse e strumenti digitali, ha acquisito una elevata autoconsapevolezza teorica (e più di un qualche riconoscimento istituzionale), ma di rado è riuscita a uscire dal circolo dei suoi addetti ai lavori, a stabilire una relazione scientifica con il *mainstream* della comunità scientifica umanistica; come ha scritto J. Unsworth:

We need (we still need) to demonstrate the usefulness of all the stuff we have digitized over the last decade and more – and usefulness not just in the form of increased access, but specifically, in what we can do with the stuff once we get it: what new questions we could ask, what old ones we could answer⁵.

Più recentemente, restringendo l'orizzonte della riflessione sugli studi letterari anche Willard McCarty ha evidenziato la questione della rilevanza critica dell'informatica letteraria

[...] literary computing is confined to providing evidence for or against what we already know or suspect. It is strongly inhibited in its capacity to surprise. Providing evidence seems justification enough, but evidence becomes increasingly

⁵ J. Unsworth, "Tool-Time, or 'Haven't We Been Here Already?': Ten Years in Humanities Computing", presentato al *Transforming Disciplines: The Humanities and Computer Science*, Washington, D.C., gennaio 18, 2003. <http://www.iath.virginia.edu/~jmu2m/carnegie-ninch.03.html>.

problematic as the volume of data exceeds the norm for critical practices formed prior to the exponential growth of online resources. As this volume increases, so does the probability of arbitrary choice, and so the ease with which any statement may be connected to any other. Good critics may do better scholarship by finding more of what they need; bad critics may be swiftly becoming worse ones more easily. The point, however, is that literary computing has thereby served only as mutely obedient handmaiden, and so done nothing much to rescue itself from its position of weakness, from which it can hardly deliver the benefits claimed for it by the faithful. It has done little to educate scholars methodologically⁶.

Quali sono i motivi di questo tutto sommato insoddisfacente panorama? È possibile individuare dove l'informatica umanistica in generale e quella letteraria in particolare hanno fallito nel cogliere il punto? Certo, gli ultimi decenni sono stati caratterizzati da voghe culturali troppo lontane dal rigore formale e dall'idea di testo come oggetto linguistico: la Teoria, quella senza aggettivi per dirla con Culler⁷, non si presta facilmente a interagire con il certosino formalismo delle strutture dati e dei linguaggi informatici – salvo poi convolare felicemente a nozze con il "decostruito" ipertesto. Per non parlare poi dell'arena vasta e multiforme degli studi culturali, che spesso di tutto si occupano meno che del testo (e però molti di questi studi non poco si gioverebbero del contributo di alcune delle recenti

⁶ W. McCarty, "Literary enquiry and experimental method: What has happened? What might?". In L. Dibattista, *Storia della Scienza e Linguistica Computazionale: Sconfinamenti Possibili*, Milano, Franco Angeli, 2009, pp. 40–41. <http://www.mccarty.org.uk/essays/McCarty,%20Literary%20enquiry%20and%20experimental%20method.pdf>.

⁷ J. D. Culler, *Teoria della letteratura : una breve introduzione*, Roma, Armando, 1999.

tendenze innovative emerse nell'ambito delle Digital Humanities e di cui parleremo più avanti).

E tuttavia, come rileva lo stesso McCarty, proprio il problema del testo, pur così al centro della riflessione teorica e metodologica delle Digital Humanities, è rimasto tutto sommato scarsamente determinato nell'ambito delle concrete applicazioni dell'informatica agli studi letterari. Il problema è che non disponiamo di teorie del testo che si possano definire in senso stretto formali. Mentre la rappresentazione (e a maggior ragione l'elaborazione) informatica è ontologicamente formale in senso stretto. La lunga storia di quel sottodominio delle Digital Humanities rubricata sotto l'etichetta di *codifica testuale* è consistita nel tentativo non soddisfacente di superare questo duplice divario concettuale. E di conseguenza gli strumenti informatici per l'analisi e l'edizione scientifica dei testi (e i relativi risultati in termini di analisi ed edizioni) hanno quasi sempre deluso le aspettative e non sono riusciti ad acquisire un sufficiente riconoscimento nell'ambito delle discipline per così dire tradizionali. Cioè, nonostante si sia consapevoli del problema teorico, la predisposizione degli strumenti di rappresentazione e analisi concreti ha finora fatto assai poco i conti con le specificità e la complessità degli oggetti e delle procedure di analisi tipiche della ricerca letteraria.

Questo è dovuto anche e soprattutto al fatto che, nonostante le ripetute affermazioni teoriche, assai sporadico e di nicchia è stato l'investimento degli stessi cultori delle Digital Humanities nella definizione di nuovi modelli e linguaggi per la rappresentazione ed elaborazione formale dei complessi oggetti culturali cui si applicano. Più comunemente si sono ereditati e applicati modelli e linguaggi elaborati dall'informatica per finalità e domini diversi.

Paradigmatico il caso del linguaggio XML. Esso ha assunto un ruolo centrale nella costruzione di linguaggi standard per la rappresentazione di dati e metadati, divenendo una sorta di esperanto digitale; in virtù della sua flessibilità, robustezza e delle sue caratteristiche sintattiche è stato ampiamente adottato per la rappresentazione dei dati in ambito umanistico. Il problema è che XML da una parte impone l'adozione di un modello di dati ad albero che non sempre si adatta alla natura strutturale degli oggetti da rappresentare, dall'altra non è in grado di rappresentare adeguatamente i numerosi e complessi livelli semantici che caratterizzano un testo letterario. Anzi, in generale possiamo dire che XML non fornisce alcuna semantica ai dati in modo computazionalmente trattabile. Il comune fraintendimento per cui si parla di "markup semantico" deriva dal fatto che i marcatori sono leggibili e che, di norma, il vocabolario dei linguaggi XML usa termini delle lingue naturali. Ma la semantica "naturale" di tale vocabolario è del tutto inaccessibile a un elaboratore XML⁸.

⁸ F. Ciotti, "La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi". In L. Perilli, D. Fiormonte (a cura di), *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlandi*, Firenze, Le Lettere, 2011

Nuove frontiere per le Digital Humanities

Se questo è il quadro, quali sono le prospettive che si possono aprire per lo sviluppo dell'informatica umanistica e di quella letteraria in particolare? Senza dubbio, consolidare e se possibile estendere i risultati acquisiti è una missione validissima e anzi irrinunciabile. Gli archivi testuali vanno preservati ed implementati, la trascrizioni ed edizioni digitali basate sui formalismi attualmente disponibili moltiplicate, gli standard mantenuti, applicati e diffusi. Ma è giunto il momento ormai di individuare nuove linee di ricerca, di esplorare le tendenze innovative che potrebbero fornire un ulteriore salto di qualità e una più ampia giustificazione scientifica (ma anche istituzionale) all'incontro tra informatica e studi umanistici. Difficile dire a priori quali direzioni saranno le più proficue: il tempo lo dirà. Tra i numerosi campi di indagine aperti, ne segnaleremmo almeno due:

- 1) *Big Data*: lo sviluppo e l'applicazione di strumenti per l'analisi automatica delle ingenti masse di risorse testuali/documentali e di dati disponibili sulla rete e non solo, attraverso la sperimentazione di metodologie e tecnologie di *text mining* e *knowledge extraction*.
- 2) *Web 3.0*: la sperimentazione dei nuovi linguaggi e modelli di dati per la rappresentazione dei livelli semantici nelle risorse informative, delle tecnologie e delle architetture che vanno rubricate sotto le etichette di Web Semantico e Linked Data, ovviamente adattandole alle specificità degli oggetti culturali.

Per quanto riguarda la prima linea di ricerca, la cui analisi esula dagli obiettivi di questo lavoro, ci limitiamo a ricordare che si tratta di un ambito di ricerca complesso ma promettente che consiste nella applicazione di tecniche e strategie di *data mining*, ovvero di metodi computazionali, su base statistico/probabilistica, per la ricerca di regolarità e schemi ricorrenti impliciti e non osservabili a priori all'interno di grandi moli di dati strutturati e non strutturati:

Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques⁹.

La ricerca di tali pattern e regolarità si basa su complessi algoritmi probabilistici, i più noti dei quali sono fondati sull'analisi probabilistica bayesiana, che studia la probabilità di eventi non quantificabili a priori (ad esempio la probabilità che in un insieme di soggetti prevalga una certa aspettativa, o che un testo sia categorizzabile in base a un dato tema prevalente). Quando questi algoritmi sono applicati a dati testuali si parla più specificamente di *text mining*. In questa direzione si sono indirizzati alcuni importanti progetti di ricerca nell'ambito delle Digital Humanities, tra cui ricordiamo un importante progetto internazionale diretto da John Unsworth, il *Monk Project*¹⁰, e le ricerche condotte presso lo *Stanford*

⁹ Gartner Group, "Data Mining Definition | Gartner", 2012, <http://www.gartner.com/it-glossary/data-mining>.

¹⁰ J. Unsworth, M. Mueller, *The MONK Project Final Report*, Settembre 2, 2009. <http://monkproject.org/MONKProjectFinalReport.pdf>.

Literary Lab diretto da Franco Moretti¹¹. Lo stesso Moretti ha teorizzato come queste tecnologie di ricerca possano essere il fondamento di un vero e proprio nuovo metodo di studio dei fenomeni letterari, che ha definito *distant reading* (giocando sulla opposizione con il *close reading* introdotto nella critica letteraria dal *New criticism*)¹².

Il Web Semantico, le ontologie e i Linked Data

In questa sede intendiamo piuttosto approfondire il discorso sulle potenzialità delle tecnologie del *Web Semantico* (WS) nelle Digital Humanities in generale e negli studi letterari in particolare. Il termine e la visione a cui esso allude sono state proposte da Tim Berners-Lee, l'inventore del Web, nel 1998¹³. L'idea consiste nell'associare alle risorse informative sul Web una descrizione formalizzata del loro significato intensionale mediante la sovrapp-

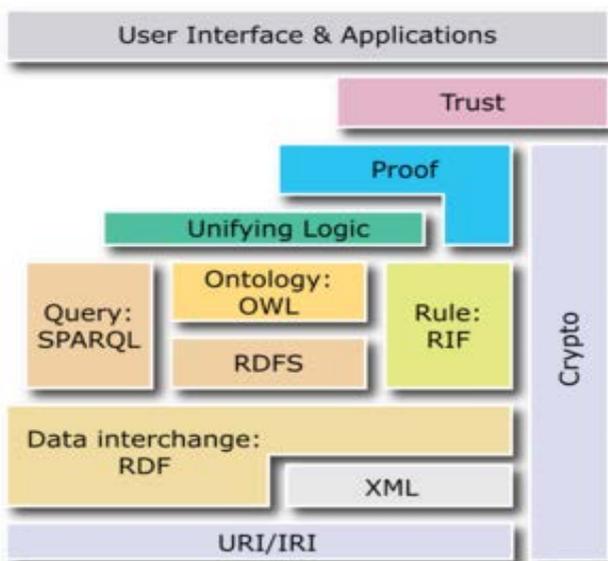
¹¹ R. Heuser, L. Le-Khac, Stanford Literary Lab, *A Quantitative literary history of 2,958 nineteenth-century British novels: the Semantic cohort method*, 2012. <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.

¹² Si vedano: F. Moretti, *Graphs, maps, trees: abstract models for a literary history*, London; New York, Verso, 2005; M. G. Kirschenbaum, "The remaking of reading: Data mining and the digital humanities". In *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, Baltimore, MD, 2007. <http://www.cs.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>

¹³ T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web". In *Scientific American*, vol. 284, fasc. 5, maggio 2001, pp. 34–43.; G. Antoniou, F. Van Harmelen, *A semantic Web primer*, Cambridge, Mass., MIT Press, 2008; E. Della Valle, I. Celino, D. Cerizza, *Semantic web: modellare e condividere per innovare*, Milano, Pearson Addison Wesley, 2008.

posizione di uno o più livelli di *metadati semantici*. Tali insiemi di metadati semantici sono espressi in formalismi che si collocano nella famiglia dei sistemi di rappresentazione della conoscenza a suo tempo sviluppati nell'ambito dell'intelligenza artificiale, e dunque possono essere elaborati automaticamente a diversi livelli di complessità: si va dalla semplice visualizzazione o consultazione per scorrimento di indici strutturati; alla interrogazione e ricerca supportata da motori inferenziali; alla classificazione e collegamento automatico; fino alla derivazione di nuove conoscenze implicite mediante inferenza logica e alla valutazione di attendibilità mediante il computo logico di asserti di fiducia.

Figura 1.



L'architettura generale del Web Semantico è comunemente raffigurata mediante un diagramma a pila che ne specifica le componenti a diversi livelli di astrazione, a partire dagli oggetti informativi a cui si applica, tecnicamente denominate *risorse*¹⁴. Il significato di questo termine è assai ampio: una risorsa informativa può essere un oggetto informativo accessibile sul Web – dal singolo documento, a sue parti, a collezioni di documenti – un oggetto reale o un oggetto astratto. Il primo problema che si pone è quello della 'identificazione' delle risorse in modo non ambiguo e indipendente dall'universo del discorso. Le URI (*Uniform Resource Identifiers*) sono i formalismi che svolgono tale ruolo: degli identificativi univoci e persistenti che permettono che una data risorsa possa essere menzionata e individuata nello spazio informativo del Web¹⁵.

Se una risorsa è identificata in modo univoco è possibile esprimere su di essa asserti che ne descrivono il contenuto sotto un qualche rispetto, esprimono ciò che un utente pensa su tale contenuto, ne specificano proprietà e relazioni. Questi asserti sono i *metadati semantici*. Affinché i metadati semantici siano utilizzabili dai

¹⁴ In effetti al momento sono stati definiti modelli architetture e linguaggi solo fino al livello delle ontologie e dei sistemi a regole. Gli strati più 'alti' del WS sono ancora avvolti da una aura quasi mistica, e molti esperti nutrono forti dubbi che potranno mai essere tradotti in qualcosa di funzionante, almeno sulla scala totalizzante prevista dal disegno di Berners-Lee. Ci sono ad esempi dei seri problemi formali e matematici che impediscono di applicare algoritmi di dimostrazione automatica a questo livello.

¹⁵ La forma più comune di URI son gli indirizzi delle pagine Web (URL) per cui esiste un consolidato protocollo di dereferenziazione, ma non sono le uniche (e peraltro la loro funzione è spuria in quanto svolgono sia il ruolo di identificatori sia quello di localizzatori).

computer, è necessario che vengano espressi in un linguaggio che sia computazionalmente trattabile. È questo il fine del *Resource Description Framework* (RDF) sviluppato presso *World Wide Web Consortium*¹⁶. RDF è un metalinguaggio dichiarativo per formalizzare asserti che esprimono proprietà e relazioni tra risorse, il cui modello di dati è basato su tre elementi:

- Risorse.
- Proprietà.
- Asserti.

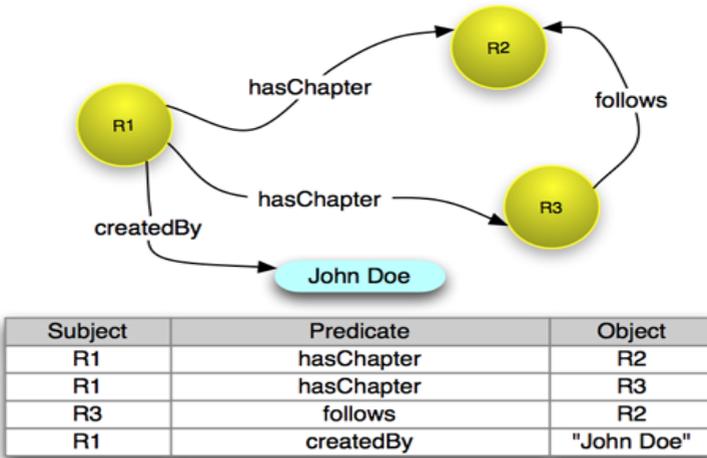
Le risorse come abbiamo visto sono tutto ciò che può essere soggetto di descrizione: pagine Web, documenti, persone, istituzioni, concetti. Le proprietà sono coppie attributo-valore associate alla risorsa. Ogni proprietà ha un significato specifico, una serie di valori leciti ed è associabile a uno o più tipi di risorsa. Proprietà e valori possono essere espressi da URI (e dunque da altre risorse) o da valori letterali (valori diretti).

Gli asserti (*statement*) sono la struttura predicativa che esprime l'associazione di una proprietà a una risorsa. Ogni asserto ha una struttura soggetto – predicato – oggetto. Un asserto specifica una

¹⁶ Il Web Consortium (W3C, <http://www.w3c.org>) è una organizzazione *no profit* che promuove e coordina lo sviluppo delle tecnologie di base per il Web in modo indipendente dai singoli attori privati a esso interessati, rilasciando standard e linee guida in regime aperto. Le specifiche di RDF sono in *Resource Description Framework (RDF): Concepts and Abstract Syntax*, K. G., Carroll J. (Ed.), W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-concepts>, che definisce il modello dei dati e la sintassi per esprimere asserti RDF.

relazione predicativa tra soggetto e oggetto (in RDF sono consentite solo relazioni binarie). Gli asserti sono anche noti come triple e gli insiemi di asserti si possono rappresentare come *grafi etichettati orientati aciclici*, come si vede nella figura seguente.

Figura 2.



RDF in quanto tale non fornisce un vocabolario predefinito e a priori di proprietà e di relazioni sotto cui sussumere e organizzare le risorse. Si tratta di un modello di dati semplice e rigoroso per specificare proprietà di risorse, qualsivoglia esse siano. In un contesto ampio ed eterogeneo come il Web possono esistere numerosi schemi e vocabolari semantici, basati su diverse concettualizzazioni di particolari domini, su diverse terminologie e lingue. In linea generale si può assumere che esistano anche concettualizzazioni mutuamente contraddittorie e/o mutevoli nel tempo. Al fine di rendere utilizzabili queste concettualizzazioni in modo

computazionale (almeno in parte) è necessario conseguire un ulteriore livello di formalizzazione: quello delle ontologie formali.

La definizione classica di questo concetto è stata fornita da Gruber¹⁷: "An ontology is an explicit specification of a conceptualization". Il termine ontologia, ereditato dalla metafisica classica dove, sin dalla sistemazione aristotelica, denotava la teoria dell'essere e delle sue categorie, è oggi adottato a denotare una ampia e diversificata classe di oggetti che vanno dai vocabolari controllati, ai thesauri fino alle ontologie formali vere e proprie. Queste, oltre a fissare una terminologia strutturata per gli enti di un dato dominio, ne fissano anche la semantica condivisa da una data comunità, in termini logico-formali:

In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application¹⁸.

Esistono numerosi linguaggi formali per specificare ontologie formali. A un primo e più semplice livello di complessità e capacità espressiva si pone RDF Schema (RDFS), che permette di definire formalmente:

¹⁷ T. R. Gruber, "A translation approach to portable ontology specifications". In Knowledge Acquisition, vol. 5, fasc. 2, 1993, p. 199.

¹⁸ T. R. Gruber, "Ontology", Encyclopedia of Database Systems, Springer-Verlag, 2009.

- Classi (o tipi) di risorse
- Classi di proprietà
- Relazioni tra classi di risorse e proprietà (Ex.: classe -> sottoclasse)
- Domini e range di proprietà

I vincoli espressi da RDFS, tuttavia, non sono sufficienti per esprimere interamente i vincoli ontologici necessari agli obiettivi del WS. Occorre un sistema per specificare le relazioni logico-semantiche (equivalenza, specificazione, generalizzazione, istanziazione, cardinalità, simmetria etc.) tra oggetti e proprietà di un medesimo schema e di schemi diversi. Ad esempio, la relazione di "autorialità" potrebbe essere indicata dalla proprietà "essere autore" dove l'autore sta in funzione di soggetto e il cui oggetto è un dato documento. In uno schema differente, al contrario, potremmo avere che il soggetto è il documento di cui si predica la proprietà "essere scritto da" che ha come oggetto un esponente della classe degli autori. Evidentemente si sta parlando dello stesso insieme di individui e relazioni (dominio), ma in modo simmetrico.

Nel contesto del WS il linguaggio deputato a conseguire questo secondo livello di formalizzazione è *Web Ontology Language* (OWL). OWL, ora giunto alla versione 2.0, può essere espresso in diverse notazioni equivalenti e ha due possibili interpretazioni semantiche: una (OWL 2 DL) basata sulla semantica modellistica di una variante

di *description logic* completa e computabile¹⁹; e una più potente (OWL 2 Full) basata su un generalizzazione del modello a grafo di RDF/S, ma non decidibile. Nelle logiche descrittive la modellizzazione ontologica è distinta in due parti: la cosiddetta *Tbox* (*terminological box*) descrive le classi e i concetti generali, specifica le loro relazioni e ne definisce proprietà; la *Abox* (*assertion box*) contiene gli asserti fattuali, che elencano gli individui del dominio, ne individuano i ruoli e indicano a quali classi e concetti definiti nella *Tbox* essi appartengano.

Questa struttura semplifica la modellazione concettuale di domini fortemente popolati, poiché gli individui possono essere descritti con una relativa autonomia dalla formalizzazione concettuale di alto livello. Le ontologie formali basate su *description logic* hanno anche il vantaggio di potere essere utilizzate da sistemi di ragionamento automatico abbastanza efficienti (quali il motore inferenziale Racer o il più antico e consolidato linguaggio Prolog), mediante i quali si possono eseguire numerosi processi di elaborazione e gestione delle basi di conoscenza:

Logic reasoning is one possible application for ontologies. It is probably helpful (i) to check consistency during ontology development, (ii) to enable semi-automatic merging of (domain) ontologies as well as (iii) to deduce hidden information contained in the ontology. These three tasks can be applied to all elements of ontologies, classes as well as instances

¹⁹ F. Baader, *The description logic handbook: theory, implementation, and applications*, Cambridge, UK; New York, Cambridge University Press, 2003.

[...] logic reasoning can fulfill different purposes in the phase of creating an ontology and in the phase of using it [...], e.g. investigating the structure of categories/ concepts, or testing if every object is used in the intended and not contradictory way. In different situations of a work process, logic reasoning can be used to avoid or to solve problems: If several persons build together an ontology, new included elements can be checked for inconsistency or redundant information can be detected²⁰.

Il progetto del WS (o Web 3.0, formulazione adottata dopo e per certi versi in alternativa al diffondersi della moda culturale/ tecnologica del Web 2.0) nella sua generalità richiede numerose e rilevanti innovazioni dal punto di vista tecnico, da quello delle competenze richieste e soprattutto da quello dei comportamenti sociali e culturali degli utenti del Web. Molti esperti e studiosi nutrono forti dubbi sul fatto che tale progetto nella sua versione più ambiziosa e universale potrà mai realizzarsi. Come accennato esistono numerosi e validi ostacoli tecnici e teorici: inconsistenza tra ontologie; incompletezza dei sistemi deduttivi per le versioni più espressive di OWL e RDFs; complessità computazionale degli algoritmi inferenziali applicati a un numero di asserti potenzialmente enorme; criticità della Assunzione di Mondo Aperto, secondo la quale è falso solo ciò che si può dimostrare esplicitamente tale, alla base delle logiche descrittive. Ma forse più rilevanti sono i dubbi circa la sua reale necessità, almeno per gli scopi e gli obiettivi

²⁰ A. Zöllner-Weber, "Ontologies and Logic Reasoning as Tools in Humanities?". In *Digital Humanities Quarterly*, vol. 3, fasc. 4, 2009. <http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html>.

per cui è stato ideato e progettato, che sarebbero assai meglio conseguiti dalle tecnologie e dai sistemi di cooperazione decentrata e sociale introdotti dal cosiddetto Web 2.0: si pensi ad esempio al meccanismo delle *folksonomie* (contrapposte alle tassonomie e ai thesauri) e al *social filtering*²¹.

Diverso il discorso relativo all'applicazione di tecnologie del WS a domini specifici e in contesti controllati e locali. In questi contesti vengono meno molte delle problematiche tecniche e vengono valorizzate le capacità di organizzazione delle conoscenze da parte di esperti pur potendo usufruire di strumenti assai più flessibili e dinamici rispetto ai tradizionali strumenti di controllo semantico dell'informazione. In questa direzione si muovono anche le recenti sperimentazioni che vanno sotto l'etichetta di *Linked Data*²².

Con questo termine ci si riferisce a un insieme di soluzioni per la pubblicazione e l'interconnessione di dati strutturati sul Web mediante tecnologie del WS. L'idea è stata introdotta ancora da Tim Berners-Lee al fine dare concretezza al progetto del WS rendendo disponibile nei suoi formalismi e protocolli la sterminata quantità di

²¹ C. Shirky, *Ontology is Overrated: Categories, Links, and Tags*, 2005. http://www.shirky.com/writings/ontology_outrated.html; K. H. Veltman, "Towards a Semantic Web for Culture". In *Journal of Digital Information*, vol. 4, fasc. 4, 2004. <http://jodi.tamu.edu/Articles/v04/i04/Veltman>.

²² T Berners-Lee, "Linked Data - Design Issues", 2006. <http://www.w3.org/DesignIssues/LinkedData.html>; C. Bizer, T. Heath, T. Berners-Lee, "Linked Data - The Story So Far", *International Journal on Semantic Web and Information Systems*, vol. 5, fasc. 3, 33 2009, pp. 1-22.; T. Heath, C. Bizer, "Linked Data: Evolving the Web into a Global Data Space", *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, fasc. 1, febbraio 2011, pp. 1-136.

dati contenuti nei database dei sistemi informativi già ora presenti sul Web. Questo deve avvenire seguendo alcuni semplici principi di base:

The term *Linked Data* refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee in his Web architecture note *Linked Data* and have become known as the *Linked Data principles*. These principles are the following:

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

The basic idea of Linked Data is to apply the general architecture of the World Wide Web to the task of sharing structured data on global scale. In order to understand these Linked Data principles, it is important to understand the architecture of the classic document Web²³.

La creazione di Linked Data, insomma, costituisce un percorso graduale dall'attuale Web come *rete di documenti* il cui contenuto di conoscenza è prodotto interamente dall'interpretazione umana a un Web come *rete di dati* che veicolano in modo formalizzato frammenti di semantica processabili da un elaboratore. I Linked Data

²³ T. Heath, C. Bizer, *op.cit.*

forniscono inoltre una soluzione al problema della identificazione condivisa dei concetti (intesi nel senso più generale possibile), grazie all'adozione condivisa di URI. Un'applicazione esemplare e molto nota di questa architettura è costituita da *GeoNames* (www.geonames.org), una ontologia (qui il termine è usato nel senso ampio di vocabolario formalizzato) che descrive circa 10 milioni di toponimi, organizzati per attributi e relazioni geografico/territoriali. Un altro *repository* di Linked Data assai noto è *DBPedia* (www.dbpedia.org) che costituisce una riformulazione formalizzata di parte del contenuto della nota enciclopedia collaborativa Wikipedia²⁴.

Studi letterari e tecnologie semantiche: verso il Web Semantico Letterario

Dato il contesto teorico e tecnico che abbiamo delineato sopra, quale convergenza ci può essere tra le tecnologie del Web Semantico e gli studi umanistici e letterari in particolare? Ha un senso parlare di un *Web Semantico Letterario*? La risposta positiva a questo quesito la forniscono implicitamente le già numerose sperimentazioni e progetti di ricerca in atto, o in via di elaborazione, in questa area. Non ci soffermeremo in questa sede su una descrizione analitica di ciascuna di esse. Intendiamo piuttosto fornire una visione a larga scala, tracciare tendenze per future esplorazioni e sperimentazioni.

²⁴ C. Bizer et al., "DBpedia - A crystallization point for the Web of Data", *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, fasc. 3, settembre 2009, pp. 154-165.

Un primo punto di convergenza consiste nella creazione e pubblicazione di Linked Data Set in ambito umanistico e letterario. A un livello molto generalista questa attività ovviamente si potrebbe avvalere di convergenze interdisciplinari anche all'interno dello stesso campo umanistico. Si pensi infatti alla possibilità di predisporre ontologie terminologiche (sullo stile di GeoNames) per:

- luoghi e spazi geografici di epoche e culture diverse;
- persone ed eventi storici;
- autori e opere;
- luoghi e spazi geografici immaginari e funzionali;
- entità e personaggi finzionali;
- temi e motivi letterari;
- temi e motivi iconologici;
- figure retoriche;
- generi e stili letterari e artistici;

Questa lista potrebbe allungarsi a piacimento, ma appare evidente come la creazione di questi repertori sotto forma di Linked Data, organizzati da un livello soggiacente di ontologie formali (a loro volta capaci di interconnettersi ad altre ontologie generali come il CIDOC CRM - <http://www.cidoc-crm.org>) permetterebbe di stabilire relazioni, interconnessioni e *mash-up* tra concetti e nozioni di diversi domini; la loro esplorazione mediante sistemi di ragionamento e dedizione automatici porterebbe a individuare e studiare in modo sistematico fenomeni altrimenti invisibili o solo

intuibili. Peraltro già la stessa costruzione di una ontologia, tanto per fare un esempio, di temi e motivi letterari richiederebbe uno sforzo di analisi concettuale che porterebbe *ipso facto* un importante contributo agli studi di critica tematica.

Tuttavia predisporre repertori, ancorché strutturati *sub specie* di ontologie e Linked Data sarebbe una piccola rivoluzione rispetto alla disponibilità dei corrispettivi repertori cartacei²⁵. Così come costruire *repositories* e biblioteche digitali di intere tradizioni letterarie in codifica XML/TEI non ha *finora* portato a grandi passi avanti nella conoscenza critico-letteraria.

Un vero e proprio "salto gestaltico" si può invece ottenere mettendo in relazione ontologie/set di Linked Data e collezioni digitali di testi. Le ontologie possono infatti fornire una semantica computabile per i documenti digitali. Questa integrazione aprirebbe la strada a indagini e ricerche su vasta scala decisamente innovative. Si pensi alla possibilità di incrociare in un dato insieme testuale e in modo esaustivo le interazioni tra determinate forme metriche e figure o tropi; oppure alla possibilità di analizzare in modo sistematico come un tema si fenomenizzi e migri tra testi di varie epoche; o ancora studiare l'evoluzione intertestuale di un personaggio e la sua relazione con generi e temi.

La realizzazione di una simile macro architettura è ovviamente assai complessa tecnicamente e altrettanto onerosa in termini di

²⁵ Si pensi a opere come: R. Ceserani, M. Domenichelli, P. Fasano, *Dizionario dei temi letterari*, Torino, UTET, 2007; A. Ferrari, *Dizionario dei luoghi letterari immaginari*, Torino, UTET, 2007.

tempo e risorse (e tuttavia, se venti anni fa qualcuno avesse detto che oggi avremmo avuto a disposizione in formato digitale, opportunamente codificate in TEI, intere tradizioni letterarie, non si sarebbe forse manifestato il medesimo scetticismo?).

Per le problematiche tecniche, una soluzione consiste nell'adozione di una strategia di annotazione semantica *multilayer* dei documenti digitali basata sul paradigma dello *stand-off markup*, invece del tradizionale e notoriamente problematico approccio di *inline markup*. La segmentazione del testo al fine di codificare i frammenti linguistico-testuali che fenomenizzano i molteplici livelli semantici incorrerebbe inevitabilmente nel problema delle "gerarchie sovrapposte" che caratterizza XML²⁶.

Nello *stand-off markup* i marcatori sono in parte o in tutto esterni rispetto alla sequenza lineare dei caratteri, eliminando alla radice la questione delle sovrapposizioni sintattiche. Naturalmente si pone il problema di esprimere formalmente il collegamento tra i metadati semantici esterni (di norma triple RDF o asserti in OWL) e i brani di testo a cui sono applicati, preservando la portabilità e la (teorica) leggibilità umana dell'insieme di documenti digitali risultante. Una elegante soluzione interamente basata su tecnologie standard W3C è stata proposta da Di Iorio e Vitali nel formalismo ontologico EARMARK²⁷.

²⁶ Su questo rimandiamo a F. Ciotti, *op.cit.* e alla bibliografia ivi contenuta.

²⁷ A. Di Iorio, S. Peroni, F. Vitali, "A Semantic Web approach to everyday overlapping markup". In *J. Am. Soc. Inf. Sci. Journal of the American Society for Information Science and Technology*, vol. 62, fasc. 9, 2011, pp. 1696–1716; A. Di Iorio, S. Peroni, F. Vitali,

The basic idea is to model EARMARK documents as collections of addressable text fragments, and to associate such text content with OWL assertions that describe structural features as well as semantic properties of (parts of) that content. As a result EARMARK allows not only documents with single hierarchies (as with XML) but also multiple overlapping hierarchies where the textual content within the markup items belongs to some hierarchies but not to others. Moreover EAMARK makes it possible to add semantic annotations to the content through assertions that may overlap with existing ones²⁸.

I frammenti di testo sono identificati come range di caratteri oppure mediante puntatori XPath e XPointer. I tre autori hanno anche sviluppato diversi strumenti per applicare il modello formale, il quale è comunque utilizzabile con qualsiasi sistema di interrogazione e ragionamento compatibile con OWL e SPARQL²⁹.

Un'ulteriore applicazione delle tecnologie del Web Semantico nello studio dei fenomeni testuali e letterari consiste nella modellizzazione ontologica delle strutture narrative. Come noto si tratta di uno dei pochi settori delle scienze del testo soggetto già in passato e indipendentemente dall'informatica, a numerosi tentativi

"Handling Markup Overlaps Using OWL". In P. Cimiano, H. Pinto (a cura di), *Knowledge Engineering and Management by the Masses*. «Lecture Notes in Computer Science», vol. 6317, Springer Berlin Heidelberg, 2010, pp. 391–400 http://dx.doi.org/10.1007/978-3-642-16438-5_29.

²⁸ A. Di Iorio, S. Peroni, F. Vitali, *op.cit.*, p. 1704.

²⁹ SPARQL (*Simple Protocol and RDF Query Language*) è un linguaggio di interrogazione per basi dati semantiche espresse in RDF sviluppato dal W3C.

di formalizzazione più o meno compiuta, sin dalle origini della narratologia negli anni '60 del secolo scorso. I modelli a suo tempo elaborati da Bremond, da Todorov e poi da Greimas³⁰ cercavano di fornire una vera e propria grammatica e semantica formale della narrazione, sul modello delle grammatiche generative della lingua a suo tempo proposte da Chomsky. Successivamente diversi autori hanno ripreso ed esteso le prime formalizzazioni inserendole nel contesto più generale della linguistica testuale (ricordiamo in particolare J. Petöfi, T. Van Dijk, W. Dressler, R. de Beaugrande³¹) o degli studi sui mondi possibili narrativi e sulla teoria logica dell'azione e le loro applicazione nell'intelligenza artificiale³². Una menzione speciale, visto il contesto, va fatta ai lavori di Giuseppe Gigliozzi, che negli anni '80 e '90 del secolo scorso sperimentò e realizzò diversi modelli computazionali dei fenomeni narrativi, come la creazione di *story grammar* per fiabe e novelle e la descrizione formale dei personaggi e dei ruoli narrativi, realizzando anche diversi programmi basati sul linguaggio Lisp, come SEB e SEBNET³³.

Questa ricca tradizione rappresenta una solida base teorica su cui innestare le più recenti e flessibili (nonché supportate dal punto

³⁰ C. Brémond, *Logica del racconto*, Milano, Bompiani, 1977; Algirdas Julien Greimas, *La semantica strutturale; ricerca di metodo*, Milano, Rizzoli Editore, 1968.

³¹ T. A. van Dijk, *Text and context*, 1977; W. U. Dressler, *Current trends in textlinguistics*, Berlin; New York, W. de Gruyter, 1978.

³² M.-L. Ryan, *Possible worlds, artificial intelligence, and narrative theory*, Bloomington, Indiana University Press, 1991; J. C. Meister, *Computing action: a narratological approach*, Berlin, Walter de Gruyter, 2003.

³³ Si vedano G. Gigliozzi, *Studi di codifica e trattamento automatico di testi*, Roma, Bulzoni, 1987; G. Gigliozzi, *Saggi di informatica umanistica*, Milano, UNICOPLI, 2008.

di viste delle implementazione software) tecnologie del Web Semantico. Si muove in questa direzione l'ontologia OWL per la descrizione formale dei personaggi letterari proposta da Zöllner-Weber; da ricordare in questa ambito anche un progetto di ricerca recentemente avviato presso il Language Technology Lab del DFKI (Deutsche Forschungszentrum für Künstliche Intelligenz) che ha come obiettivo l'individuazione e il riconoscimento dei personaggi più rilevanti nei racconti popolari attraverso l'uso combinato di ontologie e sistemi di elaborazione del linguaggio naturale. Ovviamente tali ontologie di aspetti e fenomeni intratestuali possono essere integrate con quelle intertestuali che abbiamo discusso sopra, e a loro volta connesse direttamente con i testi, fino a formare un *Semantic Web Letterario* che apre nuove prospettive per l'avanzamento del sapere teorico e critico sul patrimonio letterario.

Resta da considerare un ultimo ma non marginale aspetto: chi e con quali risorse economiche e di tempo potrebbe procedere alla costruzione di questo nostro visionario Semantic Web Letterario? La risposta è: tutta la comunità degli studi letterari. La storia e l'evoluzione del Web ha dimostrato che non solo è possibile costruire sistemi, anche di enorme complessità, attraverso un processo pubblico incrementale e cooperativo, ma che tale strategia si dimostra assai più efficiente ed efficace di quelle private, monolitiche e centralizzate. Moltissimo lavoro nella costruzione della macroarchitettura che qui proponiamo potrebbe essere condotto usando sistemi di cosiddetto *crowdsourcing* guidato, posto che esistano le opportune infrastrutture abilitanti. Il modello del *social tagging*, opportunamente corretto mediante sistemi basati su ontologie che ne

orientino e controllino l'applicazione, permetterebbe di coinvolgere studiosi esperti ma anche giovani ricercatori e cultori nel costruire e popolare le ontologie. Un simile sforzo intellettuale e tecnologico non potrebbe che essere condotto in questo modo. E il prodotto di un tale impresa non potrà che essere un bene comune, un contenuto aperto e disponibile per tutta la comunità degli studi.

Riferimenti Bibliografici

- Antoniou, Grigoris, Frank Van Harmelen, *A semantic Web primer*, Cambridge, Mass., MIT Press, 2008.
- Baader, Franz., *The description logic handbook: theory, implementation, and applications*, Cambridge, UK; New York, Cambridge University Press, 2003.
- Barnard, David T., Lou Burnard, Jean-Pierre Gaspard, Lynne A. Price, Michael Sperberg-McQueen, Giovanni Battista Varile, "Hierarchical Encoding of Text: Technical Problems and SGML Solutions". In *Computers and the Humanities*, vol. 29, fasc. 3, 1995, pp. 211–231.
- Berners-Lee, Tim, "Linked Data - Design Issues", 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, Tim, James Hendler, Ora Lassila, "The Semantic Web". In *Scientific American*, vol. 284, fasc. 5, maggio 2001, pp. 34–43.
- Bizer, Christian, Tom Heath, Tim Berners-Lee, "Linked Data - The Story So Far". In *International Journal on Semantic Web and Information Systems*, vol. 5, fasc. 3, 33 2009, pp. 1–22.
- Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann, "DBpedia - A crystallization point for the Web of Data". In *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, fasc. 3, settembre 2009, pp. 154–165.
- Brémond, Claude, *Logica del racconto*, Milano, Bompiani, 1977.
- Buzzetti, Dino, "Digital Representation and the Text Model". In *New Literary History*, vol. 33, fasc. 1, S.d., pp. 61–88.

- Ceserani, Remo, Mario Domenichelli, Pino Fasano, *Dizionario dei temi letterari*, Torino, UTET, 2007.
- Ciotti, Fabio, "La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi", Lorenzo Perilli, Domenico Fiorimonte (a cura di). In *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlandi*, Firenze, Le lettere, 2011.
- Cohen, Patricia, "Digital Keys for Unlocking the Humanities' Riches", *New York Times*, novembre 16, 2010.
http://www.nytimes.com/2010/11/17/arts/17digital.html?_r=2&emc=eta1&.
- Culler, Jonathan D., Francesco Muzzioli, Gian Paolo Castelli, *Teoria della letteratura : una breve introduzione*, Roma, Armando, 1999.
- Della Valle, Emanuele, Irene Celino, Dario Cerizza, *Semantic web : modellare e condividere per innovare*, Milano, Pearson Addison Wesley, 2008.
- Dijk, Teun A. van, *Text and context*, 1977.
- Dressler, Wolfgang U., *Current trends in textlinguistics*, Berlin; New York, W. de Gruyter, 1978.
- Ferrari, Anna, *Dizionario dei luoghi letterari immaginari*, Torino, UTET, 2007.
- Gartner Group, "Data Mining Definition | Gartner", 2012.
<http://www.gartner.com/it-glossary/data-mining>.
- Gigliozzi, Giuseppe, *Saggi di informatica umanistica*, Milano, UNICOPLI, 2008.

F. Ciotti, *Web semantico, linked data e studi letterari*

— — —, *Studi di codifica e trattamento automatico di testi*, Roma, Bulzoni, 1987.

Greimas, Algirdas Julien, *La semantica strutturale; ricerca di metodo*, Milano, Rizzoli Editore, 1968.

Gruber, Thomas R., "A translation approach to portable ontology specifications", *Knowledge Acquisition*, vol. 5, fasc. 2, 1993, pp. 199–220.

— — —, "Ontology", *Encyclopedia of Database Systems*, Springer-Verlag, 2009.

Heath, Tom, Christian Bizer, "Linked Data: Evolving the Web into a Global Data Space". In *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, fasc. 1, febbraio 2011, pp. 1–136.

Heuser, Ryan, Long Le-Khac, Stanford Literary Lab, *A Quantitative literary history of 2,958 nineteenth-century British novels: the Semantic cohort method*, 2012. <http://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.

Di Iorio, Angelo, Silvio Peroni, Fabio Vitali, "A Semantic Web approach to everyday overlapping markup". In *J. Am. Soc. Inf. Sci. Journal of the American Society for Information Science and Technology*, vol. 62, fasc. 9, 2011, pp. 1696–1716.

— — —, "Handling Markup Overlaps Using OWL". In P. Cimiano, H. Pinto (a cura di), *Knowledge Engineering and Management by the Masses*. «Lecture Notes in Computer Science», vol. 6317, Springer Berlin Heidelberg, 2010, pp. 391–400 http://dx.doi.org/10.1007/978-3-642-16438-5_29.

Kirschenbaum, M. G., "The remaking of reading: Data mining and the digital humanities". In *The National Science Foundation*

Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD, 2007. <http://www.cs.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>.

Koleva, Nikolina, Thierry Declerck, Hans-Ulrich Krieger, "An Ontology-Based Iterative Text Processing Strategy for Detecting and Recognizing Characters in Folktales", Jan Christoph Meister (a cura di). In *Digital Humanities 2012 Conference Abstracts*, 467–470, Hamburg, Hamburg University Press, 2012.

Landow, George P., *Hypertext: the convergence of contemporary critical theory and technology*, Baltimore, Johns Hopkins University Press, 1992.

McCarty, Willard, *Humanities computing*, Basingstoke [England]; New York, Palgrave Macmillan, 2005.

— — —, "Literary enquiry and experimental method: What has happened? What might?", in Liborio Dibattista, *Storia della Scienza e Linguistica Computazionale: Sconfinamenti Possibili*, 32–54, Milano, Franco Angeli, 2009. <http://www.mccarty.org.uk/essays/McCarty,%20Literary%20enquiry%20and%20experimental%20method.pdf>.

McGann, Jerome, "Marking Texts of Many Dimensions", Susan Schreibman, Ray Siemens, John Unsworth (a cura di), *A Companion to Digital Humanities*, 198–217, Oxford, Blackwell, 2004. <http://www.digitalhumanities.org/companion>.

Meister, Jan Christoph, *Computing action: a narratological approach*, Berlin, Walter de Gruyter, 2003.

F. Ciotti, *Web semantico, linked data e studi letterari*

Moretti, Franco, *Graphs, maps, trees: abstract models for a literary history*, London; New York, Verso, 2005.

Ryan, Marie-Laure, *Possible worlds, artificial intelligence, and narrative theory*, Bloomington, Indiana University Press, 1991.

Schreibman, Susan, Raymond George Siemens, John Unsworth, *A companion to digital humanities*, Malden, Mass., Blackwell Pub., 2004.

Shirky, Clay, *Ontology is Overrated: Categories, Links, and Tags*, 2005.

Universo bibliografico e semantic web

Gianfranco Crupi

Il contesto: semantic web

Nel contesto del semantic web, il termine semantico non allude alla semantica del linguaggio naturale, quanto al fatto che i dati siano trattabili da un computer, che contengano cioè informazioni valide perché la macchina possa elaborarle in modo corretto. Tuttavia, il semantic web ha una sua propria lingua, non una lingua parlata ma una lingua "inventata" per comunicare e scambiare dati e informazioni tra esseri umani e per essere letta, interpretata ed elaborata da computer. Una lingua con una sua propria grammatica, funzionale ad esprimere la natura relazionale dei dati e la loro proteiforme tipologia. Questa grammatica, denominata RDF, fornisce la struttura logica per gestire ed esprimere le relazioni tra le informazioni, ispirandosi ai principi della logica dei predicati, secondo la quale le informazioni sono esprimibili mediante *asserzioni* (o 'statement'), costituite da un essenziale modello sintagmatico tripartito (*tripla*):

- un *soggetto*, vale a dire una qualsiasi *risorsa*, non necessariamente accessibile via web, che identifica la "cosa" descritta: *documenti* (leggibili da esseri umani) o *oggetti* (leggibili da una macchina);
- un *predicato*, cioè una *proprietà* specifica della risorsa o una relazione utilizzata per descriverla, identificata da un *nome*;
- un *oggetto*, detto *valore*.

Inoltre, secondo la grammatica RDF, ogni frase o asserzione descrive il rapporto tra due entità – per esempio, tra un'opera e il suo autore ("Giuseppe Verdi ha composto *La Traviata*") – o tra un'entità e le annotazioni testuali che la connotano (ad esempio, la parola "*La Traviata*" e le parole che indicano la data e il luogo della sua prima rappresentazione: "6 marzo 1853", "Venezia", "Teatro La Fenice"). E tuttavia, come si è già detto, fatta eccezione per le annotazioni testuali, ogni elemento di un'asserzione RDF è rappresentato, nella sua grammatica, non da parole della lingua parlata ma da stringhe di caratteri precedute dal prefisso "http://", che identificano in modo univoco qualsiasi risorsa (URI, *Uniform Resource Identifier*): da un indirizzo web a uno di posta elettronica, da un documento a un servizio, da un file a un programma, ecc. Nella lingua del semantic web, l'URI consente inoltre l'utilizzo dell'oggetto identificato in contesti differenti da quello d'origine e a prescindere dalla sua espressione testuale.

«Un URI può essere classificato come URL o come URN.

- Un URL (Uniform Resource Locator) è un URI che, oltre a identificare una risorsa, fornisce mezzi per agire su o per ottenere una rappresentazione della risorsa descrivendo il suo meccanismo di accesso primario o la sua "ubicazione" ("location") in una rete.
- Per esempio, l'URL <http://www.wikipedia.org> è un URI che identifica una risorsa (l'home page di Wikipedia) e lascia intendere che una rappresentazione di tale risorsa (il codice HTML della versione corrente di tale home page) è ottenibile via http da un host di rete chiamato www.wikipedia.org.
- Un URN (Uniform Resource Name) è un URI che identifica una risorsa mediante un "nome" in un particolare dominio di nomi ("namespace"). Un URN può essere usato per parlare di una risorsa senza lasciar intendere la sua ubicazione o come ottenerne una rappresentazione. Per esempio l'URN `urn:isbn 0-395-36341-1` è un URI che consente di individuare univocamente un libro mediante il suo nome `0-395-36341-1` nel namespace dei codici ISBN, ma non suggerisce dove e come possiamo ottenere una copia di tale libro»¹.

¹ "Uniform Resource Identifier". In *Wikipedia. L'enciclopedia libera*, http://it.wikipedia.org/wiki/Uniform_Resource_Identifier, 04.12.2003; ultima modifica 04.08.2012.

Ciascuna asserzione RDF è esprimibile da un *grafo* formato da *nodi* e *archi* che rappresentano le risorse, le loro proprietà e i rispettivi valori. Questo modello a grafi, per essere pubblicato, viene codificato in formati di *serializzazione*², tramite i quali diventa possibile per le macchine elaborare il modello e comprendere il significato delle descrizioni delle risorse.

Figura 1: Article indexed in AGRIS database.

The screenshot shows a detailed record for an article in the AGRIS database. On the left, there is a list of metadata fields including Alternative Title, Date of publication, AGRIS Categories, AGROVOC terms, AGROVOC French terms, AGROVOC Spanish terms, Language, Notes, Journal Title, ISSN, and Vol. No. The main title of the article is 'Study on the mortality and competition of three dominant grasshoppers (Orthoptera: acrididae) in the steppe'. On the right side, there are 'Related AGRIS Results' listed as bullet points, and a section for 'Other related searches by title' and 'Related searches by author/s'. A 'beta OpenAgris' logo is visible at the top right of the record area.

Go to AGRIS search	beta OpenAgris
	<i>Try it!</i>
Acta Agrestia Sinica (Sep. 2008)	
Study on the mortality and competition of three dominant grasshoppers (Orthoptera: acrididae) in the steppe	
Lu Hui; Han Jianguo	
Alternative Title	典型草原三种蝗虫种群死亡率和竞争的研究
Date of publication	Sep. 2008
AGRIS Categories	Animal ecology
AGROVOC	Acrididae; Mortality; Steppes
English terms	
AGROVOC French terms	Acrididae; Mortalite; Steppe
AGROVOC Spanish terms	Acrididae; Mortalidad; Estepas
Language	Chinese
Notes	16 ref.
Journal Title	Acta Agrestia Sinica
ISSN	1007-0435
Vol. No.	v.16(5) p.480-484

Related AGRIS Results:

- **Predictions of species interactions from consumer-resource theory: experimental tests with grasshoppers and plants** by Ritchie, M.E. (Utah State Univ., Logan (USA), Dept. of Fisheries and Wildlife); Tilman, D. (1993) in English
- **Intra- and interspecific competition in adults of two abundant grasshoppers (Orthoptera: Acrididae) from a sandhills grassland** by Joern, A.; Klucas, G. (Apr 1993) in English
- **Toward a general model of rangeland grasshopper (Orthoptera: Acrididae) phenology in the steppe region of Montana** by Kemp, W.P. (USDA, ARS, Rangeland Insect Laboratory, Bozeman, MT); Dennis, B. (Dec 1991) in English

> **Other related searches by title**

> **Related searches by author/s**

powered by Google™

▼ About the Title

[Study on the Mortality and Competition of Three Dominant Grasshoppers \(Orthoptera:Acrididae\) in the Steppe.](#) LU Hui,HAN Jian-guo(Institute of Grassland ...

² Per "serializzazione" s'intende il processo di conversione di una struttura di dati in un formato che può memorizzato e quindi rigenerato nello stesso o in un altro ambiente informatico.

Tabella 1: Statements about Article as "pseudo-triples".³

Subject	Predicate	Object
Article	has the Title	典型草原三种蝗虫种群死亡率的研究
Article	has an Author	Lu Hui
Article	has a Topic	Acrididae (grasshoppers)

Tabella 2: Triplifying the description...³

Subject	Predicate	Object
agris:CN2009002389	has the Title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	has an Author	agris-author:luhui
agris:CN2009002389	has a Topic	Acrididae (grasshoppers)

The article is identified by the URI <http://agris.fao.org/resource/CN2009002389>.
Let's abbreviate this as [agris:CN2009002389](#).

³ L'esempio, adattato e modificato, è tratto da T. Baker, "Designing data for the open world of the web (Slides)". In *JLIS.it, Online First* (Giugno/June 2012): Art. #6308. DOI: 10.4403/jlis.it-6308.

Tabella 3: *Triplifying the description...3*

Subject	Predicate	Object
agris:CN2009002389	dct:title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	dct:creator	agris-author:luhui
agris-author:luhui	foaf:name	Lu Hui
agris:CN2009002389	dct:subject	agrovoc:c_4416
agrovoc:c_4416	skos:prefLabel	Acrididae (en)
agrovoc:c_4416	skos:prefLabel	蝗科 (zh)

A property for the preferred label of a concept has been defined
 In the Simple Knowledge Organization System (SKOS) vocabulary.

Più precisamente, gli identificatori utilizzati da RDF sono gli *URI references* (URIfref), ovvero identificatori formati da un URI, cui è aggiunto un suffisso con caratteri Unicode, che consentono così di esprimere e definire le relazioni tra qualsiasi "cosa". Sebbene gli oggetti, che rappresentano i valori associati ai predicati, possano essere espressi come stringhe di caratteri (dette *literals*), l'utilizzo di URIfref consente alle applicazioni di distinguere le proprietà che potrebbero essere identificate con lo stesso nome *literal* e che possono a loro volta essere trattate come risorse, dando la possibilità di associare loro informazioni aggiuntive. «Un indirizzo URI – grazie al modo in cui è formato – porta dunque in sé, almeno in modo implicito, una citazione. Gli indirizzi di tipo URI usati per le

proprietà e per le classi portano il lettore a definizioni documentate in maniera ufficiale. Così è il Web stesso che fornisce alla lingua dei dati il suo dizionario»⁴.

Inoltre, la natura linguistica che informa l'intero sistema rende possibile la costruzione di frasi multiple, relative allo stesso soggetto, o di frasi che, in base al principio di *inferenza*, ne generano di nuove, dando vita a un reticolo di asserzioni, e quindi a un complesso di relazioni (secondo un modello desunto dalla logica dei database relazionali), che amplia la rete semantica degli ambiti di appartenenza dei dati, espressi nelle singole asserzioni. L'assimilazione del principio della combinatorietà, secondo il quale – com'è noto – un numero *limitato* di unità minori si possono combinare tra loro per formare un numero *illimitato* di unità maggiori, rende dunque possibile la produzione di messaggi che contengono livelli sempre più alti di complessità relazionale e, al contempo, di granularità, relativa al dominio di appartenenza dei singoli oggetti.

Sono dunque i sintagmi – segmenti di frasi che possono essere formati da una o più parole, costituenti le *asserzioni* – e le funzioni sintattiche che essi assumono nella frase, a sollecitare e favorire l'integrazione di dati provenienti da fonti differenti, generando pertanto nuovi collegamenti fra nodi, grazie a regole ontologiche basate sul significato delle proprietà e delle risorse descritte.

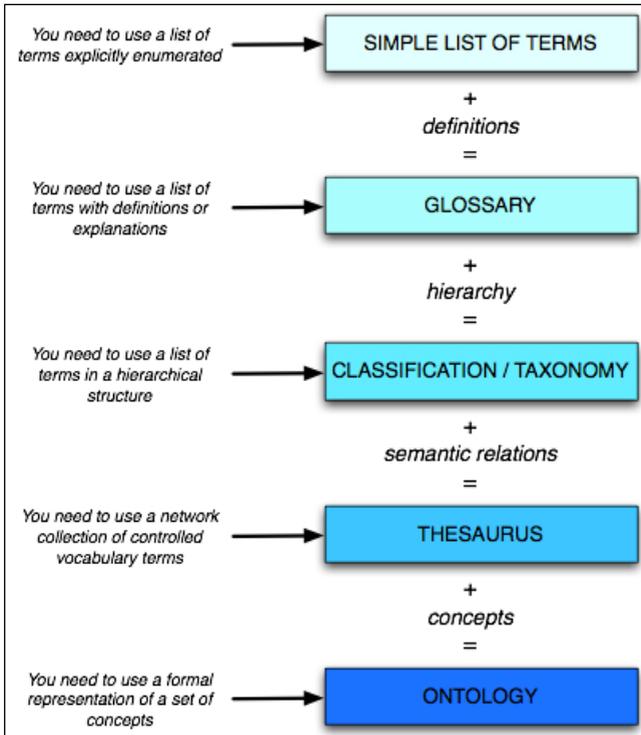
⁴ T. Baker, "Designing data for the open world of the web". In *JLIS.it, Online First* (Giugno/June 2012): Art. #6308. DOI: 10.4403/jlis.it-6308.

Del pari, i ruoli semantici che i costituenti sintagmatici, in rapporto tra loro, assumono all'interno dell'evento rappresentato dalla frase, non si identificano con le funzioni sintattiche e dipendono invece dall'azione evocata dal verbo. Va da sé che il potenziale informativo dei sintagmi è nel rapporto tra la predicazione contenuta nel messaggio, veicolato dalla frase, e l'entità cui la predicazione si riferisce.

Ebbene perché questo semplice ma articolato sistema linguistico funzioni correttamente «è necessario utilizzare un'infrastruttura tecnologica in cui i concetti siano univocamente identificati e in cui agenti software riconoscano questi oggetti e realizzino associazioni ed equivalenze tra essi»⁵. Questa infrastruttura tecnologica è costituita da una serie di strumenti condivisi di controllo terminologico e di disambiguazione semantica, che permettono di descrivere univocamente dati e di esprimere la loro semantica formale: si tratta sostanzialmente di linguaggi, metalinguaggi, vocabolari controllati e ontologie, la cui tipologia può essere così rappresentata:

⁵ M. Guerrini, P. Possemato, "Linked data: un nuovo alfabeto del web semantico". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5463. DOI: 10.4403/jlis.it-5463.

Tabella 4.⁶



Ci riferiamo innanzitutto a quella famiglia di linguaggi di rappresentazione della conoscenza, finalizzati alla creazione di ontologie e destinati ad essere elaborati e interpretati dalle macchine, denominata Web Ontology Language (OWL), sviluppata dal W3C

⁶ Schema tratto da: M.-V. Leroi, "Linked Heritage: a collaborative terminology management platform for a network of multilingual thesauri and controlled vocabularies". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5471. DOI: 10.4403/jlis.it-5471

(World Wide Web Consortium)⁷. Con OWL è possibile definire ed esprimere *ontologie*, cioè strutture logiche in cui si codifica la semantica di uno specifico dominio del sapere⁸, che esplicitano quel che sappiamo di esso, tramite classi, relazioni fra classi e individui appartenenti a classi; una conoscenza elaborabile automaticamente, che consente di implementare i processi inferenziali e deduttivi. Insomma, scopo di OWL è la descrizione di basi di conoscenza, l'elaborazione di deduzioni su di esse e la loro integrazione con i contenuti delle pagine web, fornendo un linguaggio che consente una maggiore e migliore integrazione di dati tra comunità che descrivono il loro dominio.

Com'è noto, le ontologie hanno una struttura concettuale simile a quella dei thesauri, solo che esse possono prevedere un numero di relazioni superiori, generando così una complessa rete di connessioni fra concetti, che può essere visualizzata anche graficamente. Inoltre, loro caratteristica specifica è la capacità di esprimere i concetti in maniera non ambigua e quindi con un alto livello di precisione semantica. «L'opera di armonizzazione delle

⁷ L'acronimo OWL, al posto del più corretto WOL, fu adottato dal Gruppo di lavoro del W3C perché più facilmente memorizzabile, anche per la sua omofonia con il termine che in lingua inglese indica il 'gufo'.

⁸ «La modellizzazione della realtà in forme analizzabili secondo regole fisse è chiamata anche *ontologia formale*. Nel nostro contesto il termine ha chiaramente un senso più applicativo, e alcuni filosofi storcono il naso quando viene utilizzata la stessa parola per indicarlo. Tuttavia, tra i due significati ci sono alcune affinità: se riusciamo a modellare in modo più fedele la struttura della realtà, sapremo anche costruire dei sistemi più efficaci di organizzazione della conoscenza» (C. Gnoli, V. Marino, L. Rosati, *Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione per il Web*, Milano, HOPS – Tecniche nuove, 2006, pp. 44-45).

ontologie e degli schemi descrittivi viene affidata ad agenti software che, disponendo di una rappresentazione della conoscenza e di regole di deduzione espresse con un linguaggio interoperabile, operano per armonizzare conoscenze diverse»⁹.

C'è poi la famiglia dei linguaggi formali, destinati alla rappresentazione di thesauri, schemi di classificazione, tassonomie, soggettari e altre tipologie di vocabolari controllati strutturati¹⁰, che è costituita dal Simple Knowledge Organization System (SKOS)¹¹. Si tratta, anche in questo caso, di un'applicazione RDF, che offre la possibilità di definire relazioni semantiche tra i concetti e che può essere usata come formato d'interscambio¹². La sua duttilità consente l'interazione con altri strumenti e vocabolari usati nel web semantico,

⁹ O. Signore, "Semantic web: il futuro è già qui?". In *JEK POT The joint between Economy Knowledge Pathos Organization and Technology*, 10° Knowledge Management Forum, 24-25 Novembre 2005, <http://www.w3c.it/papers/km10.pdf>.

¹⁰ Nell'ambito del knowledge management, un vocabolario controllato è una raccolta di termini che possono essere utilizzati per l'indicizzazione dei documenti, ed è, per l'appunto, un vocabolario costituito solo da espressioni controllate. Un thesaurus è – secondo la definizione dello standard ISO – un «vocabolario di un linguaggio d'indicizzazione controllato», che rende esplicite le relazioni semantiche tra concetti, come quella gerarchica e quella associativa (ISO International Organization for Standardization, 1986, *ISO 2788: Linee guida per la costruzione e lo sviluppo di thesauri monolingue*, ISO).

¹¹ SKOS è un *data model* sviluppato dal W3C Semantic Web Deployment Working Group (SWD WG) e adottato da molte Biblioteche nazionali per i propri vocabolari controllati.

¹² Un esempio di thesaurus, costruito secondo il framework di SKOS, è quello creato a supporto delle attività di indicizzazione nell'ambito archivistico del Regno Unito, UKAT (United Kingdom Archival Thesaurus): <http://www.ukat.org.uk>.

come ad esempio GeoNames¹³ (un database geografico che fornisce gli strumenti per tradurre i luoghi geografici in dati che li rappresentano: latitudine, longitudine, altezza, popolazione, codice postale, ecc.) o Friend of a Friend (FOAF)¹⁴, che, facendo propria la logica e la filosofia dei social network, codifica in formato standard dati personali nonché le relazioni e i contatti che le persone intrattengono e stabiliscono all'interno di gruppi e comunità.

Tra le tecnologie convergenti verso il semantic web c'è poi quella costituita dalle *topic map*¹⁵, che, come RDF, è «una tecnologia basata sul concetto di identità. Essa utilizza simboli che rappresentano "cose" identificabili sul web (anche se spesso non recuperabili in esso) per poter fare affermazioni su di esse»¹⁶. Le *topic map* «forniscono funzionalità combinate di indici, glossari e thesauri, creando così potenti meccanismi di navigazione tra vaste collezioni di risorse digitali interconnesse tra loro, laddove questo tipo di interconnessione non debba essere necessariamente fisica bensì solamente concettuale. Ciò grazie al salto di astrazione che viene compiuto: queste mappe non si pongono allo stesso livello del documento o della risorsa, ma sono sovrapponibili, si posizionano a un livello superiore e formano un comune sovrastato semantico agli oggetti cui fanno riferimento e che vengono "mappati". In questo

¹³ <http://www.geonames.org>

¹⁴ <http://www.foaf-project.org>

¹⁵ *Topic Maps* è uno standard ISO: ISO/IEC 13250:2003 (Information technology - SGML applications - Topic maps); rivisto poi e modificato, è divenuto ISO/IEC 13250-2:2006.

¹⁶ "Topic Maps". In *Wikipedia. L'enciclopedia libera*, http://it.wikipedia.org/wiki/Topic_Maps, 26.04.2007; ultima modifica: 10 marzo 2012.

modo più mappe possono essere applicate alle stesse informazioni oppure la stessa mappa può essere applicata a gruppi di informazioni diverse, permettendo un elevato livello di flessibilità e personalizzazione. La struttura proposta è di tipo reticolare e a più livelli, utilizzando uno schema che si presta molto di più al sistema della ricerca scientifica e alle modalità di organizzazione del pensiero, superando i limiti delle strutture lineari e ad albero imposte dai formati di memorizzazione dei supporti informatici»¹⁷.

Il funzionamento logico dei linked data

Il paradigma che è alla base dei linked data¹⁸ è quello del web tradizionale, il web ipertestuale o dei documenti, incentrato, come si sa, su pochi ma efficaci standard: l'HTML, in quanto linguaggio di markup e formato di impaginazione, formattazione e visualizzazione; l'HTTP, il protocollo universale per la trasmissione di

¹⁷ F. Meschini, "Le mappe topiche. Come imparai a non preoccuparmi e ad amare i metadati", *Bollettino AIB. Rivista italiana di biblioteconomia e scienza dell'informazione*. Vol. 45, 2005(1), p. 62, <http://bollettino.aib.it/article/view/5389/5155>.

¹⁸ La denominazione di linked data si riferisce a un «set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee in his Web architecture note *Linked Data* and have become known as the *Linked Data principles*» (T. Heath, C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, San Rafael, Morgan & Claypool, 2011, p. 7. DOI 10.2200/S00334ED1V01Y201102WBE001). I principi formalizzati da T. Berners-Lee sono i seguenti: 1. Use URIs as names for things; 2. Use HTTP URIs, so that people can look up those names; 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL); 4. Include links to other URIs, so that they can discover more things.

informazioni in modalità ipertestuale; l'URI, il sistema di identificazione unica e universale.

Ebbene, questa "semplice" architettura logica è quella su cui poggiano i principi fondamentali per la pubblicazione nel web di dati strutturati e per la loro condivisione: l'utilizzo di URI, per identificare non solo documenti web e contenuti digitali, ma anche oggetti del mondo reale e concetti astratti, anche perché essi funzionano come mezzo di accesso alle informazioni che descrivono l'entità identificata; l'adozione di HTTP URI, che consentano agli URI di essere *dereferenziati*, tramite il protocollo HTTP, in una descrizione dell'oggetto identificato o del concetto astratto; e, infine, l'uso di un meccanismo standard per specificare l'esistenza e il significato delle connessioni tra gli elementi descritti nei dati, offerto da RDF, che, attraverso dichiarazioni di relazioni tra le "cose" del mondo (persone, luoghi o concetti astratti) espresse in link qualificati, fornisce appunto un modo flessibile per descriverle, per indicare le relazioni che esse intrattengono con altre "cose" e per dichiarare esplicitamente la natura di tali relazioni.

La tecnologia dei linked data è dunque vincolata al modello RDF, non solo perché esso fornisce su scala globale l'identificazione unica di entità ma anche perché consente l'uso in parallelo di schemi differenti per la rappresentazione dei dati. Attraverso RDF e l'utilizzo degli URI, come identificatori universali delle entità, i linked data mettono in naturale relazione entità provenienti da diverse e sempre nuove fonti di dati.

Questo processo di messa in relazione e, soprattutto, di integrazione di dati provenienti da fonti diverse è reso possibile dal riferimento a vocabolari condivisi (che rendono perciò le definizioni

delle parole recuperabili) e dal fatto che termini di differenti vocabolari siano collegati tra di loro mediante link tra i vocabolari stessi, sulla cui scelta non esiste, da parte degli editori di dati, alcun vincolo pregiudiziale. E questo in base al presupposto che i dati siano correttamente strutturati (*conditio sine qua non* della loro riutilizzabilità) e che siano autodescrittivi, vale a dire che se un'applicazione trova dati descritti con un vocabolario sconosciuto, l'applicazione può dereferenziare l'URI che identifica i termini del vocabolario, al fine di trovare la loro definizione, consentendo così alle applicazioni client di scoprire tutte le pertinenti meta-informazioni, necessarie per integrare i dati provenienti da fonti diverse. Insomma, la riusabilità dei dati è sollecitata dalla natura autodescrittiva dei linked data, nel senso che ogni proprietà utilizzata per descrivere la relazione tra due "cose" è essa stessa descritta utilizzando lo stesso formato di dati che descrive i dati¹⁹.

La dereferenzialità, cui si è fatto cenno, significa che i client possono cercare l'URI utilizzando il protocollo HTTP e recuperare in tal modo una descrizione della risorsa (sia essa un documento HTML, un oggetto del mondo reale o un concetto astratto) che viene identificata dall'URI; le descrizioni delle risorse, che sono destinate all'elaborazione da parte delle macchine, sono rappresentate come dati RDF. Tuttavia, quando gli URI identificano "cose" del mondo reale, onde evitare rischi di ambiguità, confondendo le "cose" con i documenti che le descrivono, è di prassi l'utilizzo di URI differenti,

¹⁹ T. Hodson, *British Library Data Model: Overview*, <http://talisisystems.com/2011/07/british-library-data-model-overview>

distinguendo così in modo coerente le asserzioni su una "cosa" dal documento che la descrive.

Nell'articolazione linguistica del modello RDF, la logica dei link è quella di rompere l'autoreferenzialità dei dati, moltiplicando le relazioni con altre fonti di dati che, ad esempio, forniscano informazioni di contesto circa l'identità di una persona o il luogo in cui egli vive. Inoltre, il fatto che essi puntino a URI differenti, per riferirsi a una medesima "cosa" del mondo reale o a un medesimo concetto astratto, consente di documentare e di esprimere la polisemia e la pluralità di punti di vista che intorno ad essi esistono.

La promessa del web, modellato sulla logica dei linked data, non è solo quella di consentire alle applicazioni client di scoprire nuove fonti di dati, seguendo i link RDF in fase di esecuzione, ma anche di aiutarli a integrare i dati provenienti da queste fonti²⁰; infatti, informazioni desunte da fonti diverse possono essere facilmente combinate attraverso la fusione in un unico grafo delle due serie di *triple*.

È importante ricordare che RDF non è un formato di dati, ma un modello di dati per descrivere le risorse in forma di *triple* costituite da soggetto, predicato e oggetto. Tuttavia, fornendo RDF solo un generico, astratto modello di dati per la descrizione delle risorse, l'integrazione, dal punto di vista semantico, avviene soprattutto attraverso operazioni di mappatura, utilizzando tassonomie, lessici e ontologie, espressi – come si è detto - in linguaggi e schemi di rappresentazione della conoscenza come OWL, SKOS e RDFS (il linguaggio RDF Vocabulary Description, meglio noto come RDF

²⁰ Cfr. K. Coyle, *Linked Data Tools. Connecting on the Web*, Chicago, ALA Editions, 2012.

Schema). Essi soddisfano appunto l'esigenza di esprimere tassonomie, thesauri, soggettari (SKOS) e di fornire vocabolari per descrivere modelli concettuali, in termini di classi e delle loro proprietà, nonché i rapporti di sussunzione tra termini (RDFS, OWL).

I linked data e il mondo delle biblioteche

Linked data si configura dunque come un'applicazione dei principi del web finalizzata a un nuovo e più flessibile paradigma editoriale dei dati. Il risultato è uno spazio di dati globale - il Web di dati, appunto - basato su standard aperti e costituito da un numero incalcolabile di asserzioni RDF provenienti dalle fonti più disparate e che coprono i più svariati argomenti. È qui che risiede la fortuna che la tecnologia dei linked data sta riscuotendo in ogni settore dell'interazione sociale nel web e, più specificatamente, nell'ambito del *cultural heritage* e della comunicazione scientifica.

In particolare, le biblioteche scoprono oggi l'opportunità di integrare le informazioni strutturate dei loro cataloghi con informazioni provenienti da altri cataloghi e da fonti terze (come, ad esempio, DBpedia²¹), e a rendere più facilmente accessibili i loro dati basandosi su standard web. Il problema è che il catalogo della biblioteca, per essere visibile all'utente, deve cessare di essere un'entità staccata, un database separato, un *silo* isolato dal web, ma deve essere integrato in esso, da esso interrogabile, capace di parlare

²¹ DBpedia è un progetto collaborativo per l'estrazione e il riutilizzo di informazioni semanticamente strutturate da Wikipedia e per rendere queste informazioni disponibili sul web e riutilizzabili da parte di software e applicazioni.

e di capire la lingua del web, vale a dire la lingua degli utenti che nel web "vivono" e operano, come fosse il loro habitat naturale, e dove si sono candidati nuovi soggetti, concorrenti a popolare l'universo della mediazione informativa e che competono autorevolmente a ridisegnare la geografia dei saperi e i luoghi dell'accesso alla conoscenza. La trasformazione del catalogo in un sistema integrato con la tecnologia che la gente utilizza per la ricerca e per la creazione di nuove idee, è possibile se esso esce dalla dimensione autoreferenziale che, per certi versi, lo ha finora contraddistinto, per incontrare le esigenze di utenti, che non siano necessariamente solo quelli elettivi del catalogo tradizionale, e che utilizzano normalmente il web come loro principale fonte di informazione. Ciò comporta l'elaborazione di un diverso sistema d'uso e di fruizione dei dati bibliografici, in grado di rispondere con maggiore aderenza al funzionamento del web e alle regole di una socialità espansa, che ha fatto propria la filosofia dell'accesso aperto alle fonti della conoscenza e, soprattutto, ai dati, alla loro proteiforme varietà, ai dati che sono di per sé relazioni, che sono la connessione strutturale tra le "cose" e la cui combinazione genera sempre nuova conoscenza.

Parola chiave di tale processo è "interoperabilità", non solo tecnologica ma soprattutto semantica, culturale – verrebbe da dire; quella che nasce dall'incontro di comunità e etnie digitali diverse, con i loro linguaggi, le loro tradizioni, il loro differente modo di classificare e rappresentare le "cose" del mondo. È un concetto – quello di interoperabilità – che il mondo delle biblioteche conosce bene perché lo ha teorizzato e praticato negli ultimi decenni. Oggi, il problema è quello di «come rendere utilizzabili i dati bibliografici nel web, usando sia il potere di calcolo che abbiamo a disposizione che

la potenza combinatoria del web»²². La tecnologia offerta dai *linked data* è un'opportunità di straordinaria importanza, sebbene non l'unica possibile. «Ma non possiamo entrare nell'ambiente informativo, ricco e dinamico, del Web del XXI secolo con dati basati sui principi del XIX secolo»²³. Interoperabilità significa allora – in questa specifica accezione - rendere accessibili e disponibili i dati, in modo tale da essere elaborabili da una macchina per consentire la loro integrazione e il loro riutilizzo in applicazioni differenti. Le esperienze pilota della Bibliothèque nationale de France²⁴, della Library of Congress²⁵, della Bayerische Staatsbibliothek²⁶, della

²² K. Coyle, "Library Linked Data: un'evoluzione". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5443. DOI: 10.4403/jlis.it-5443.

²³ *Ibid.*

²⁴ La Bibliothèque nationale de France con il progetto data.bnf.fr dà accesso, attraverso un'unica interfaccia web, ai documenti digitali in suo possesso e ai dati descrittivi provenienti dai suoi diversi cataloghi. L'interoperabilità tra le diverse fonti catalografiche e documentarie della BNF e tra esse e quelle provenienti da set di dati esterni è assicurata dall'adozione degli standard del semantic web e dallo loro espressione secondo il modello concettuale di FRBR. Cfr. BNF, *Présentation générale du projet data.bnf.fr* (15.11.2011), <http://data.bnf.fr/docs/databnf-presentation.pdf>; R. Wenz, "Linked open data for new library services: the example of data.bnf.fr". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5509. DOI: 10.4403/jlis.it-5509.

²⁵ La Library of Congress ha avviato un progetto per rendere disponibili, secondo il formato linked data e senza restrizioni d'uso, i suoi vocabolari controllati, tra cui un primo nucleo di classi tratte dal sistema di classificazione LCC (Library of Congress Classification). Cfr. Library of Congress, *LC Linked Data Service. Authorities and Vocabularies*, <http://id.loc.gov>; K. Ford, "Library of Congress Classification as Linked Data". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5465. DOI: 10.4403/jlis.it-5465.

²⁶ Presso la Deutsche Nationalbibliothek e l'Hochschulbibliotheksentrums des Landes Nordrhein-Westfalen, il centro di servizio bibliotecario North-Rhine-Westphalian (*hbz*), è stato istituito un servizio di *linked open data*, (Culturegraph,

British Library²⁷ e di OCLC²⁸ indicano con grande evidenza che il mondo delle biblioteche (ma anche quello degli archivi e dei musei) entra nel mondo del semantic web, immettendo in esso una consolidata tradizione di teorie e di pratiche fondate sul controllo bibliografico e sul controllo di autorità dei dati, sulla sensibilità e sulla capacità di gestire le informazioni, catalogando il sapere, creando nuove connessioni semantiche tra i documenti, fornendo dunque loro valore aggiunto attraverso la struttura sindetica dei cataloghi, gli strumenti indicali, i linguaggi di indicizzazione semantica e di classificazione. Una sensibilità, che oggi si traduce nella progettazione dei nuovi contesti digitali e degli spazi logici di interazione tra gli utenti e l'universo documentario e dei servizi, per consentire un accesso intuitivo ai contenuti e un loro facile recupero. Ecco perché è fondamentale che i dati strutturati e controllati delle biblioteche siano presenti in rete e accessibili con nuovi strumenti,

questo è il suo nome) che genera un identificatore unico e specifico per tutti i tipi di risorse possedute dalle biblioteche tedesche con l'obiettivo di creare un catalogo di metadati aperti.

²⁷ La British Library sta sviluppando una versione della British National Bibliography (BNB) sotto forma di open linked data e secondo un [modello concettuale](#) che è stato efficacemente rappresentato in forma grafica. Il progetto riguarda nella sua fase iniziale le monografie e le pubblicazioni seriali. Cfr. T. Hodson, *British Library Data Model: Overview*, cit.; British Library, Free data services, <http://www.bl.uk/bibliographic/datafree.html>.

²⁸ OCLC ha di recente reso disponibili oltre un milione di linked data (corrispondenti a circa 80 milioni di triple) relativi a pubblicazioni tra le più presenti su WorldCat (il catalogo unico di OCLC), selezionate in base al numero di localizzazioni (almeno 250) di ciascun documento. L'intero [progetto](#) è illustrato da un [video](#), *Linked Data for Libraries*, che è anche una sintetica e utile introduzione alla tecnologia dei linked data.

compatibili con le tecnologie e gli standard del web. I linked data daranno vita a nuovi servizi sostanzialmente basati sul bagaglio di conoscenze e di pratiche che sono parte integrante della tradizione di biblioteche, archivi, musei, preoccupati da sempre di convertire l'informazione in dati e metadati di qualità.

L'opportunità offerta da questo nuovo metodo di pubblicare dati sul web, costituito dai linked data, comporterà, se pienamente messo a frutto, una radicale trasformazione del rapporto tra utente e universo bibliografico:

- l'integrazione dei propri dati con quelli di altri soggetti istituzionali non solo aumenta, infatti, il loro potenziale informativo ma li rende più completi, più usabili e riusabili in contesti anche assai diversi da quelli d'origine;
- la chiarezza esplicativa del linguaggio usato nel web rende meno oscuro, e quindi più comprensibile all'utente, il linguaggio della biblioteca e degli strumenti semantici di classificazione e organizzazione della conoscenza da essa adottati;
- l'aggregazione e il collegamento con altre risorse web, anche se strutturate secondo standard differenti, consente di estendere illimitatamente le informazioni di contesto relative a ciascun dato;
- l'incontro con altri segmenti del web accresce il numero degli strumenti di controllo terminologico a disposizione, rendendo più precise e pertinenti le fonti informative, la cui riconosciuta autorevolezza è criterio distintivo fondamentale per conferire legittimità e validità ai dati;

- far uscire fuori dal "deep web" dati locali e renderli aperti, universalmente accessibili, significa offrire un'opportunità democratica di visibilità a culture minoritarie;
- l'integrazione dei dati catalografici all'interno del web semantico comporta, di riflesso, un arricchimento degli stessi cataloghi e la possibilità di offrire nuovi servizi basati sulla tecnologia e sul linguaggio del web;
- inoltre, «il recente accordo – noto come schema.org – tra i maggiori motori di ricerca (Google, Yahoo, Bing e il russo Yandex) per la codifica nel linguaggio RDF dei dati all'interno delle normali pagine HTML (HTML5) può (o deve) essere anche per le biblioteche una interessante opportunità. Grazie a questa codifica – che si presenta come un'estensione molto semplice dei tag HTML delle pagine web ma che si fonda sul linguaggio RDF – i motori di ricerca sono in grado di "capire" la struttura e la tipologia di un determinato documento. Con una codifica basata su schema.org i nostri cataloghi, grazie ai dati strutturati che contengono, potranno essere "oggetti semantici" interpretabili dai maggiori motori di ricerca»²⁹.

²⁹ G. Bergamin, A. Lucarelli, "Il Nuovo soggettario come servizio per il mondo dei linked data". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5474. DOI: 10.4403/jlis.it-5474.

Com'è noto, la qualità di una biblioteca non si misura tanto dalla quantità di documenti posseduti, quanto dalla capacità di strutturare e di modellizzare i dati, di renderli accessibili, conservando al contempo la stratificazione dei contesti, la relazione tra quello nuovo che si crea e il contesto di origine, oltre che con tutti gli altri documenti con cui essi stringono relazioni semantiche tacite o palesi. Essa cioè deve saper ricostruire le relazioni logiche e genetiche tra i documenti, rendendoli tuttavia disponibili a nuove ricomposizioni semantiche, lasciate dinamicamente a totale giudizio e scelta dell'utente; deve, insomma, saper esibire i contesti multipli a cui i documenti fanno riferimento. Di qui la necessità di uno sforzo – già in atto – per rendere identificabili in modo univoco i propri dati nel contesto del web e per renderli disponibili a essere leggibili, interpretabili e usabili dalle macchine.

La comunità internazionale dei bibliotecari si è già attivamente mossa dando vita – come si è detto - a importanti progetti di trasformazione e di adeguamento dei cataloghi. L'esperienza citata della Bibliothèque nationale de France ci induce a pensare che il catalogo del prossimo futuro e i dati bibliografici avranno una forma e una funzione ben diverse da quelle attuali: un *catalogo-enciclopedia* che espone tutte le relazioni possibili tra i dati in suo possesso e quelli recuperati da altre fonti, che diventa esso stesso conoscenza elaborata, strumento primario di *reference*. E un analogo sforzo è stato messo in atto dagli organismi nazionali e internazionali (IFLA in testa) per cercare di tradurre in linked data schemi bibliografici e

di classificazione come ISBD³⁰, FRBR³¹, RDA³², DDC³³, LC Classification³⁴, il Nuovo soggettario italiano³⁵. Nell'un caso e nell'altro, si tratta di delicate operazioni che gravano sull'architettura logica di complessi sistemi documentari e normativi, e che pongono

³⁰ L'IFLA ISBD Review Group si è di recente attivato con l'obiettivo di «improving the portability of bibliographic data in the semantic web and consequently the interoperability of the ISBD standard in connection with other content standards» (IFLA, *ISBD: International standard bibliographic description*, Consolidated ed., Berlin et al., De Gruyter Saur, 2011, p. 1).

³¹ Tra gli obiettivi al centro di interesse del FRBR Review Group c'è quello di promuovere gli standard dell'IFLA e di partecipare alla creazione di namespace per tutti gli standard bibliografici (ISBD, FRBR, FRAD, FRASD inclusi) «and in connection with this promote and position the IFLA standards and models in the semantic web» (*Action Plan for 2012*, <http://www.ifla.org/en/node/1959>); cfr. P. Riva, "FRBR Review Group initiatives and the world of linked data". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5482. DOI: 10.4403/jlis.it-5482.

³² Tra gli obiettivi dichiarati del Joint Steering Committee for Development of RDA (Resource Description and Access), il nuovo standard catalografico che rimpiazza l'AACR2, c'è quello di rendere i dati «adaptable to new and emerging database structures» (Joint Steering Committee for Development of RDA, *Resource Description and Access*, <http://www.rda-jsc.org/rdafaq.html>).

³³ Già dal 2009, OCLC si è impegnata a pubblicare la Classificazione Decimale Dewey come vocabolario controllato di linked data; l'iniziativa è in corso di realizzazione. Cfr. J. S. Mitchell, M. Panzer, "Dewey linked data: Making connections with old friends and new acquaintances". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5467. DOI: 10.4403/jlis.it-5467.

³⁴ Cfr. nota 22.

³⁵ Da novembre 2010 il Nuovo soggettario, curato dalla Biblioteca Nazionale Centrale di Firenze, ha reso disponibili i suoi metadati nel formato RDF/SKOS, con l'obiettivo di favorire la loro "usabilità" nel mondo dei Linked data. Cfr. G. Bergamin, A. Lucarelli, *Il Nuovo soggettario come servizio per il mondo dei linked data*, cit.

rilevanti problematiche di coerenza sistemica, inerenti soprattutto alle scelte linguistiche e alla gestione dei diritti relativi ai dati.

Le prime, volte a salvaguardare e a garantire il multi-linguismo e la diversità linguistica e culturale con azioni (com'è nel caso di ISBD) finalizzate all'adozione di URI opachi, espressi in cifre, poiché «le dichiarazioni [degli URI] contengono una serie di informazioni importanti, come il nome del metadato, l'etichetta, la definizione, le note di ampliamento dell'informazione o della sua applicazione, la filiazione (se è proprietà o subproprietà), lo stato di accettazione, etc. [...] Usando un URI opaco, è possibile riunire tutte le disposizioni in differenti lingue sotto un unico URI, semplicemente identificando la lingua in cui si desidera l'informazione [...] Rendere questi URI opachi [facilita] inoltre l'uso da parte di comunità linguistiche diverse dall'inglese e l'accesso alle traduzioni di queste ontologie in altre lingue, senza la necessità di creare URI indipendenti»³⁶.

Per quanto riguarda le seconde - le scelte, cioè, relative alla gestione dei diritti sui dati -, esse sono condizionate sia dal livello di controllo che l'editore dei dati stessi vuole esercitare su di essi, sia dalla loro intrinseca natura e tipologia. Ma, più in generale, esse pongono un problema di interoperabilità giuridica, relativa all'integrazione di dati provenienti da fonti diverse (pubbliche e private), raggiungibile, naturalmente, tramite l'evoluzione e l'armonizzazione dei contesti giuridici nazionali nel campo dei dati

³⁶ E. Escolano Rodríguez, "L'adattamento di ISBD al web semantico dei dati bibliografici espressi in linked data". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5484. DOI: 10.4403/jlis.it-5484.

pubblici, e l'adozione di idonei schemi di licenza d'uso, che, ad oggi, possono essere ricondotti a due classi:

«*Open licence* – This allows any use of the data, especially including commercial use, sometimes with restrictions about attribution and misuse.

Not-open licence – This restricts uses to non-commercial only, with similar requirements for attribution and misuse.

With both classes there are a range of standard licences, e.g. those provided by Creative Commons and GNU, and the option of a specific organisational licence»³⁷.

Significativa in tal senso l'esperienza tedesca, nelle cui reti bibliotecarie bavaresi e di Berlino-Brandeburgo si è acceso un interessante dibattito sugli aspetti giuridici degli *open data* e, in particolare, sulla pubblicità totale o parziale dei record bibliografici sotto forma di *open* o *linked open data*, che ha portato alla decisione di pubblicare record quanto più completi possibili, fatta eccezione per gli URL collegati agli indici resi disponibili da fornitori di servizi commerciali, che non possono perciò essere pubblicati per ragioni di copyright. Sebbene, ci sia chi, all'interno di quella comunità bibliotecaria, sostiene invece che non vadano gratuitamente resi disponibili quei campi di un record che hanno significativi costi di

³⁷ G. McKenna, "Linked Heritage Experience in Linking Heritage Information", in *JLIS.it, Online First* (Giugno/June 2012): Art. #6304. DOI: 10.4403/jlis.it-6304.

produzione, come ad esempio quelli relativi all'indicizzazione semantica dei documenti³⁸.

Il web semantico è – come abbiamo detto - un ambiente informativo molto eterogeneo, naturalmente portato all'ibridazione e alla contaminazione dei contenuti e dei dati provenienti da fonti diverse. Questo, se da una parte rappresenta un limite per il mondo delle biblioteche, attente alla qualità e autorevolezza delle fonti informative e a difendere la legittimità della propria terminologia e degli strumenti linguistici di controllo formale dei dati; dall'altra, l'integrazione di dati selezionati, strutturati e omogenei con quelli spesso destrutturati, provenienti da ambienti informativi molto disomogenei (ricerca scientifica, commercio, pubblica amministrazione, comunità *crowd-sourced*, ecc.), costituisce una sfida che le biblioteche non possono non raccogliere «pena la morte dei cataloghi, abbandonati dagli utenti a favore di altri strumenti di reperimento dell'informazione, come i motori di ricerca»³⁹. Sebbene, anche a fronte della crescita esponenziale delle risorse digitali, è innegabile che accanto al perseguimento degli obiettivi del progetto *Linked Open Data*⁴⁰ (rendere, cioè, accessibili i dati in formati non

³⁸ G. Messmer, "Linking library metadata to the web. The German experiences". In *JLIS.it, Online First* (Giugno/June 2012): Art. #5507. DOI: 10.4403/jlis.it-5507.

³⁹ M. Guerrini, "Saluto e introduzione ai lavori". In *Seminar "Global Interoperability and Linked Data in Libraries"*, Firenze 18-19 giugno 2012. Pubblicato su *JLIS.it, Online First* (Giugno/June 2012): Art. #6307. DOI: 10.4403/jlis.it-6307.

⁴⁰ *Linked Open Data* (LOD) promuove la disponibilità dei dati di fonte pubblica e privata, istituzionale e commerciale, in modalità tale da essere aperti a ogni genere di applicazione e perciò riusabili in contesti diversi da quelli d'origine. L'*open data* è

proprietari, collegandoli ad altri set di dati che servano a disambiguare il contenuto e a fornire loro un contesto semantico) vada garantita la certificazione di qualità dei dati e delle loro fonti, soprattutto per quel che riguarda i requisiti di integrità e autenticità⁴¹. L'impiego di sistemi di identificazione persistente rappresenta per certo (e di questo ne è consapevole il "movimento" linked data) la soluzione più convincente, quella che può garantire l'"usabilità" dei dati sul lungo termine e la loro effettiva interoperabilità; il che comporta la scelta di idonee soluzioni tecnologiche e l'adozione di autorevoli sistemi di certificazione e di accreditamento (anche a livello non istituzionale) delle comunità di utenti che li adottano.

Tuttavia, perché gli *open linked data* entrino a far parte, in modo diffuso, nella sfera operativa degli strumenti scientifici e della pratica professionale dei bibliotecari è necessario che, com'è stato osservato, questa nuova e diversa metodologia venga vissuta come un'opportunità per le biblioteche e non come un ostacolo alla loro crescita: «Linked Data becomes more powerful the more of it there

l'infrastruttura di cui i *linked data* hanno bisogno per poter creare la rete di inferenze tra i dati sparsi nel web. Pubblica amministrazione, istruzione, infrastrutture e ricerca sono solo alcuni delle potenziali aree in cui l'accesso libero ai dati può portare giovamenti e aprire nuove opportunità. Cfr. F. Bauer, M. Kaltenböck, *Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers*, Vienna, DGS, 2012, <http://www.semantic-web.at/LOD-TheEssentials.pdf>.

⁴¹ M. Lunghi, C. Cirinnà, E. Bellini, Trust and Persistence for Internet resources. In *JLIS.it, Online First* (Giugno/June 2012): Art. #5494. DOI: 10.4403/jlis.it-5494.

is. Until there is enough linking between collections and imaginative uses of data collections there is a danger librarians will see linked data as simply another metadata standard, rather than the powerful discovery tool it will underpin»⁴².

*Stop hugging your data*⁴³, dunque; questo il titolo di una conferenza di Tim Berners-Lee che qualche anno fa invitava a rendere accessibili i propri dati e a farli uscire dai *silos*, in cui sono conservati e sigillati, e non a costruirne di migliori e più efficienti. Ora sappiamo che quell'invito aveva un senso. I dati acquistano valore di conoscenza quando sono interconnessi con altri dati, quando la loro interconnessione produce deflagranti effetti di rete. E la rivoluzione copernicana dei linked data consiste proprio nel fatto che il *link*, strumento di collegamento tra documenti nel web tradizionale, acquista, nel contesto del semantic web, un ruolo semantico primario, una funzione predicativa che dà significato ai dati stessi, poiché rappresenta ed esprime i differenti tipi di relazione che essi possono intrattenere. Una rivoluzione che comporta – come si è visto - la frantumazione dell'informazione in singole componenti atomiche, in unità parcellizzate, ricombinabili con funzioni e per fini differenti. Questi principi, che costituiscono il paradigma dei linked data, applicati nel mondo del *cultural heritage* modificano (e alcuni

⁴² G. Byrne, L. Goddard, "The Strongest Link: Libraries and Linked Data". In *D-Lib Magazine*. Vol. 16, n. 11/12 (November/December 2010), doi:10.1045/november2010-byrne. <http://www.dlib.org/dlib/november10/byrne/11byrne.html>.

⁴³ T. Berners-Lee, "Stop hugging your data", TED Conferences 2009. In *YouTube*, http://www.youtube.com/watch?v=OM6XICm_qo.

esemplari esperienze lo dimostrano) i processi cognitivi che hanno finora governato la nostra relazione con l'universo bibliografico e con gli strumenti che storicamente hanno mediato il rapporto tra lettore e conoscenza (cataloghi, record, apparati indicali, ecc.). Alla base, l'idea che una visione del mondo sia possibile solo a partire dalla consapevolezza che la conoscenza è un processo dinamico di composizione e disarticolazione continua di ciò che scopriamo e sappiamo del mondo.

La narrazione in rete dei patrimoni culturali digitalizzati¹

Emiliano Ilardi

Quali ipertesti per la rete?

Dall'ottobre del 1996, un seminario finanziato dal National Research Council (NRC) aveva riunito rappresentanti dell'industria del divertimento, del cinema e dei videogiochi, insieme a membri del dipartimento della Difesa e universitari. Tutti i partecipanti la cui attività aveva a che fare con i nuovi media pensavano che il miglior modo di favorire l'immersione sperimentale nei media elettronici non fosse l'ipertesto, bensì le buone vecchie tecniche dello storytelling. Alex Seiden, rappresentante della Light and Magic, una società che si occupa di effetti speciali, spiegò: «Non ho mai visto una pagina web che avesse un grande impatto emotivo. Il racconto è la forma di espressione fondamentale dell'umanità: il romanzo il

¹ Parte di questo saggio è tratta da E. Ilardi, "Digital storytelling e istituzioni culturali". In D. Capaldi, E. Ilardi, G. Ragone (ed.), *I cantieri della memoria. Digital Heritage e istituzioni culturali*. Napoli, Liguori, 2011.

teatro, il film [...] Sono forme che determinano la nostra esperienza culturale»².

Affermazione sorprendente questa, tenendo conto che proviene da un informatico esperto di videogiochi ed effetti speciali. I nuovi linguaggi digitali, se vogliono essere efficacemente immersivi, devono ricorrere agli strumenti narrativi delle vecchie forme espressive basate sui media tipografici ed elettrici: romanzo, cinema, radio, televisione. Senza una rimediazione con essi, internet non è altro che un contenitore, un semplice database magari con l'aggiunta di un motore di ricerca. Ed effettivamente le Pagine Gialle o l'elenco telefonico (anche se on line) difficilmente riescono ad avere un grande impatto emotivo sul pubblico. Né attivano i suoi processi di memoria, tema che è al centro di questa riflessione. Eppure se ci andiamo a rileggere le prime pionieristiche riflessioni sulla comunicazione mediata dal computer le cose sarebbero dovute andare diversamente.

Secondo i primi teorici della rete e degli ipertesti³ il web sarebbe dovuto essere la perfetta realizzazione di quella visione del mondo che dalla fine degli anni Settanta andava sotto il nome di condizione postmoderna e che vedeva il superamento della modernità proprio nella crisi della funzione sociale e identitaria delle narrazioni. L'uomo nuovo nato dall'ibridazione dei linguaggi del consumo con

² C. Salmon, *Storytelling: La fabbrica delle storie*. Roma, Fazi, 2008, p. 56.

³ Cfr. ad esempio il classico saggio di G. P. Landow, *Iper testo. Il futuro della scrittura*. Bologna, Baskerville, 1993; ma anche L. Manovich, *Il linguaggio dei nuovi media*. Milano, Olivares, 2002.

le nuove tecnologie prima elettriche e poi digitali non avrebbe più avuto la necessità di doversi riconoscere in una Storia e in una memoria di lunga durata, in narrazioni coerenti, lineari, logico-causali. La rete di internet con la sua struttura reticolare e ipertestuale rappresentava perfettamente il suo nuovo habitat in cui avrebbe costruito in totale libertà i propri percorsi in base al desiderio o all'emozione del momento. Qualcosa che, con una ventina d'anni di anticipo, avevano suggerito i cosiddetti romanzieri postmoderni (da Calvino a Pynchon, dal Nouveau Roman a Barthes) con la loro intenzione di creare lettori consapevoli svelando loro i meccanismi narrativi che stanno dietro le storie.

Ecco spiegata in ambito accademico la fortuna che ha goduto fin dall'inizio il concetto di ipertesto: era visto come uno strumento liberatorio in quanto obbligava a una presa di coscienza delle sue regole di funzionamento. Prima dell'avvento delle interfacce, infatti, l'utente era costretto a conoscere i linguaggi di programmazione e gli algoritmi che stavano alla base dell'architettura dei sistemi operativi. Era il sogno di un individuo ormai perfettamente disincantato e cosciente di tutti i meccanismi che facevano funzionare la rete e che, grazie a questa nuova consapevolezza acquisita, poteva liberarsi delle logiche immediate e passivamente immersive dei media di massa⁴ e divenire padrone dell'intero meccanismo di estensione del sé.

⁴ J. D. Bolter, R. Grusin, *Remediation. Competizione e integrazione tra media vecchi e nuovi*. Milano, Guerini, 2002.

Ma internet ha tradito tutti questi intellettuali postmoderni, poststrutturalisti o decostruzionisti e soprattutto nella sua versione 2.0 ha saputo ri-mediarsi con le tecniche immersive dei media di massa, dei linguaggi della pubblicità e del consumo. Per ragioni economiche? Per colpa dei vili speculatori che mercificano inesorabilmente le istanze liberatorie dei primi rivoluzionari pionieri? In parte sì, ma non solo. Come afferma l'esperto di videogiochi citato all'inizio di questo paragrafo, la narrazione è la forma di espressione fondamentale dell'umanità. Siamo sicuri che anche oggi, in piena era digitale, gli individui frammentati nella rete non abbiano più bisogno di narrazioni forti e condivise che organizzino le loro esperienze passate in comunità, nazioni, identità di gruppo? Rifiutano coscientemente di inserire la propria memoria individuale di soggetti in narrazioni collettive? L'avvento di internet li obbliga a conoscere l'architettura della rete e li rende padroni consapevoli della produzione di simboli, narrazioni o immaginario? Sembra piuttosto il contrario. Proprio con la diffusione dei nuovi media e in misura maggiore rispetto ai periodi precedenti, il racconto e la narrazione sembrano ormai divenuti gli unici strumenti per strutturare, governare ma anche persuadere e affascinare le soggettività ormai liquefatte nel mondo di flusso e globalizzato in cui ci troviamo a vivere. Dal marketing alla pubblicità (il passaggio dal *brand image* al *brand story*), dai media alla politica fino ad arrivare alla religione, la lotta è tra chi racconta la storia più convincente, quella che crea i più forti legami identitari e di memoria. Si pensi ad esempio all'uso politico della Storia che hanno fatto i partiti italiani negli ultimi anni. Incapaci di immaginare il futuro hanno utilizzato il

passato per produrre appartenenza e militanza. Lo ha fatto Berlusconi con la sua martellante narrazione anticomunista, la Lega reinventando *ex novo* la Storia di un territorio, l'area giustizialista (dall'IDV al Movimento 5 Stelle) con il suo culto dei caduti di mafia e camorra, l'estrema sinistra con l'antifascismo militante (gli iscritti all'Associazione Nazionale Partigiani sono in continuo e costante aumento negli ultimi anni, soprattutto tra i giovani) e l'estrema destra con il suo fascismo del terzo millennio (incarnato ad esempio da Casa Pound). Solo il Partito Democratico, con la sua ossessione per le strutture organizzative leggere e reticolari, si è rifiutato di sfruttare politicamente la Storia e così lascia che siano gli altri a utilizzarla per i loro scopi. Non solo: permette ai suoi avversari, senza mai reagire, di raccontare, in termini immancabilmente negativi, la stessa Storia della sinistra. Chi oggi si rifiuta di produrre narrazioni e spettacolarizzazioni è inesorabilmente vittima delle narrazioni degli altri. Uso politico della Storia vuol dire costruire un serbatoio limitato di modelli narrativi capaci, all'occorrenza, di conferire un surplus di legittimità simbolica ai progetti per il futuro. Mettere il progetto alla fine di un processo permette ai singoli individui di potervi inserire le loro storie ed esperienze individuali. In America questa è una tecnica di comunicazione politica presente da sempre. Le guerre in Afghanistan e in Iraq sono state legittimate simbolicamente facendo un massiccio ricorso alla Storia americana: dalla Guerra di Indipendenza contro gli inglesi fino alla Seconda guerra mondiale la Storia degli Stati Uniti è diventata Storia del *free world*. Emblematiche sono state le campagne elettorali di George W. Bush, soprattutto la prima che lo vedeva in forte svantaggio rispetto

ad Al Gore. Mentre quest'ultimo snocciolava in maniera quasi autistica dati (sulla sanità, l'istruzione, l'economia ecc.), Bush raccontava la sua storia personale di cristiano rinato (*born again*) rifacendosi al modello dei risvegli religiosi tipici di tutta la Storia americana dal XIX secolo in poi. Lo stesso Obama, l'uomo del BlackBerry, si è presentato come il compimento della lunga Storia dell'emancipazione dei neri; cita continuamente Lincoln e addirittura, in certi casi, ne ha assunto lo stile e le movenze.

Il problema insomma non è che la Storia sia finita, o che le persone non abbiano più bisogno di inserire la propria memoria individuale in narrazioni collettive affinché acquisisca un senso compiuto e coerente. Il problema è che oggi si sono moltiplicati a dismisura gli agenti produttori di queste narrazioni. Neanche le cosiddette grandi narrazioni sembrano scomparse basti pensare a quella culturalista (oriente contro occidente), a quella religiosa (cristianesimo contro islam) oppure alle narrazioni basate sulle retoriche della paura o della catastrofe. Dall'invenzione dei media elettrici e poi dei media digitali i custodi del tempo si sono moltiplicati. Gli stessi mass media, messi all'angolo dai media digitali su molti versanti della comunicazione, sembrano ritrovare, nella ricomposizione di una memoria condivisa, un nuovo ruolo, una rifunzionalizzazione favorita dal fatto che la frammentarietà del web ha difficoltà a ricostruire una narrazione storica effettivamente collettiva (si pensi alle innumerevoli fiction storiche che sono state prodotte negli ultimi anni).

La narrazione e la rete

Ed è proprio di fronte a questa perdurante resistenza dei mass media che anche il cosiddetto web 2.0 ha cominciato ad integrare tecniche, software, percorsi di navigazione molto più basati sull'immediatezza narrativa dei media elettrici che sull'ipermediazione ipertestuale.

Ciò che ha fatto esplodere il web in tutto il mondo non è l'ipertesto ma una forma particolare di ipertesto quello fortemente personalizzato che produce reti sociali ma basate sull'individuo e sulla sua storia personale⁵: social network e social media come YouTube, Facebook, Flickr, Wikipedia si configurano sempre più come ambiti di condivisione di narrazioni e non tanto di semplice scambio di dati. Incastonano le informazioni in frame narrative in cui ciascuno può raccontare la propria esperienza e condividerla con gli altri. Quando pubblichiamo gli eventi della nostra vita o altrui su Facebook o YouTube, Flickr, Twitter siamo obbligati a seguire le norme di stesura che essi hanno preparato. Raccontiamo le nostre storie seguendo regole narrative imposte dall'esterno. Iper testi sì ma inseriti in frame narrative che non a caso danno vita a community e non esiste comunità anche la più digitale che non condivida un minimo di memoria comune. D'altronde, basta pensare alla cura con cui i cosiddetti hacker (la prima community on line che è apparsa) custodiscono il racconto delle loro origini, le loro radici comuni che affondano nella controcultura californiana degli anni Settanta. Essi

⁵ M. Castells, *La nascita della società in rete*. Milano, EGEA, 2002.

sono divisi su tutto per quanto riguarda il presente e il futuro ma diventano comunità proprio nel momento in cui si identificano in una storia condivisa. Ed è riconoscendosi in questo racconto collettivo che personaggi carismatici come Steve Jobs o Linux Torvalds hanno basato i loro successi e creato un popolo di fedeli (io direi quasi militanti) sparso in tutto il mondo che diffonde il loro credo (e i loro prodotti).

In questa immane battaglia che si sta svolgendo per il controllo delle sorgenti narrative collettive che ruolo giocano le vecchie istituzioni culturali (scuole, università, musei, istituti culturali, archivi)? Soprattutto, obbligate a confrontarsi con lo spazio dei flussi, sono in grado di mantenere una loro specificità, di produrre i loro specifici punti di vista e narrazioni che possano competere nel mercato globale delle storie? Da un'analisi svolta presso il Centro di Ricerca 'Digilab' dell'Università di Roma "La Sapienza" su 200 siti web di istituzioni culturali italiane la risposta a quest'ultima domanda è negativa⁶. L'impressione è che il comparto dei beni culturali e l'umanistica in generale anche quando si aprono alla rete lo facciano nella logica inefficace dell'ipertesto e del database. Ovviamente il problema della digitalizzazione, dell'archiviazione, della metadattazione, della creazione di ontologie e tassonomie è fondamentale ma il passaggio alla rete non può ridursi a questo. Manca la dimensione della comunicazione, della narrazione, della spettacolarizzazione, della ri-mediazione con le altre forme

⁶ D. Capaldi, E. Ilardi, G. Ragone, *I cantieri della memoria. Digital Heritage e istituzioni culturali*. Napoli, Liguori, 2011.

espressive. Riducendo al minimo la messa in scena spettacolare dei prodotti digitali, il rischio è di consegnarli ai racconti spesso deteriori dei media di massa e del mercato. Ci sono cioè pochissimi tentativi di provare a raccontare in rete i beni culturali e non solo semplicemente di digitalizzarli ed esporli in una pagina web.

I siti web delle istituzioni culturali continuano a preferire, nella presentazione in rete dei loro patrimoni, il formato dell'elenco invece di provare a sperimentare formati più narrativi. La maggior parte dei siti internet delle istituzioni culturali europee e soprattutto italiane, salvo casi eccezionali, si limita a costruire meri contenitori, copie digitali non dinamiche del catalogo cartaceo, liste di oggetti e riproduzioni accompagnate da didascalie fuori luogo o incomprensibili. In realtà è ciò che accade molto spesso anche nelle istituzioni "fisiche". Il bene culturale è nella maggior parte dei casi una monade, collocata nel vuoto spinto, da contemplare a distanza, irriducibile a qualsiasi narrazione o manipolazione che non provenga dall'alto, nella speranza che un giorno, chissà, possa riapparire l'aura di benjaminiana memoria. Alle potenti narrazioni storiche degli agenti del mercato e dei media le istituzioni culturali rispondono con degli elenchi convinte che, semplicemente permettendo l'accesso a tutti, il loro sito si popolerà di utenti.

Il digital storytelling per i beni culturali

Che ci piaccia o meno le nuove generazioni di nativi digitali rifiutano di essere spettatori passivi di una trasmissione di dati unidirezionale. Vogliono intervenire, commentare, condividere, produrre, creare. Il successo del web 2.0 sta proprio nell'apertura della rete alla creatività degli utenti e alla loro esigenza di rendere pubbliche le loro storie o opinioni. Anche lo stesso atto di consumare oggi non si riduce più all'identificazione in un logo o una marca ma nel consumo stesso della propria storia individuale e delle storie individuali degli altri pubblicate e messe in scena in qualche blog o social network. Il primo obiettivo di un istituzione culturale in rete deve essere quindi quello di inserirsi in maniera originale in questo processo. Proprio come fanno i grandi social network deve saper costruire i propri specifici paradigmi o contesti narrativi, norme di stesura, scenografie digitali in cui l'utente possa inserire la propria storia o esteriorizzare la sua creatività e punto di vista. Il digital storytelling applicato ai beni culturali ha tre obiettivi:

- a) riattivare i processi di memoria saldando memoria individuale e collettiva;
- b) dare identità al sito internet dell'istituzione, differenziandolo dagli altri. Questo processo oggi non può essere attivato senza inventare strategie che mettano al centro l'attività dei soggetti, le loro capacità creative: la dimensione ludica dei videogames, la dimensione cooperativa-narrativa dei wiki e dei blog, la dimensione costruttivista e didattica dell'e-learning, quella informativa del social tagging (in cui ontologie e tassonomie

non sono più imposte dall'alto). Non deve scomparire la funzione pedagogica delle istituzioni ma essa deve essere presentata non più come un processo top down (visto che tale modalità la si sperimenta già a scuola), ma come la possibilità per gli utenti di migliorare la propria storia inserendola in un contesto più ampio. Solo in questo modo ogni "luogo" del patrimonio virtualizzato può essere valorizzato divenendo il punto di incontro tra le storie prodotte dagli utenti e quelle prodotte dall'istituzione stessa.

- c) valorizzare i saperi di nicchia, i luoghi oscuri dei patrimoni e degli archivi. C'è una sorta di cupo pessimismo che attraversa gli studi umanistici, soprattutto quelli letterari: l'idea che lo studio della letteratura non importi più niente a nessuno. In realtà questa è un'impressione basata su un unico campione di riferimento: quello degli studenti universitari. Nel momento in cui l'umanistica si apre alla rete, si sta aprendo a un pubblico indifferenziato di milioni di persone. Nessuno può sapere quanti potrebbero essere gli interessati. Le strategie immersive di comunicazione servono ad attirare quel pubblico di potenziali utenti che hanno aspettative e livelli di cultura differenti. E' la strategia della "coda lunga" di Chris Anderson⁷. Più un database è grande, più ha bisogno di essere spettacolarizzato per permetterne sia l'orientamento e la navigazione da parte degli utenti, sia valorizzazione di ogni

⁷ C. Anderson, *La coda lunga. Da un mercato di massa a una massa di mercati*. Torino, Codice, 2011.

nicchia (la coda). Ecco perché ogni archivio digitalizzato e poi caricato in rete dovrebbe concentrarsi innanzitutto su una vetrina in cui i pezzi forti del patrimonio vengano spettacolarizzati in modo da diventare una specie di faro o di autostrada che guidi ogni tipologia di utente verso le strade secondarie. D'altronde ciò che di specifico offre la rete è la possibilità di presentare nello stesso contesto diverse narrazioni dello stesso oggetto. L'istituzione può cioè fornire all'utente diversi livelli di messa in scena del proprio patrimonio (dalla divulgazione più spettacolare fino al trattato accademico). Lo storytelling può essere anche scientifico non solo puro entertainment. E non è detto che l'utente, una volta catturata la sua attenzione attraverso narrazioni più spettacolari, non voglia, proprio come nei videogiochi, saltare di livello e passare a una narrazione più complessa.

The book as a medium: interesting phenomena to be researched with the help of digital humanities approaches

Christoph Bläsi

Introduction

In this paper, the range of possible cases for the application of Digital humanities approaches around the book (typically the texts conveyed by them are in the focus) is extended to ones that are less philological than rather medium-centric, covering e.g. 'material' aspects of the communication. In addition to the research problem areas dwelled on in the following in more detail, book-related Semantic Web / Linked Open Data applications or indeed certain aspects of digital and transmedia storytelling could be objects of digital humanities approaches in book studies research – beyond a narrow concept of the philological, which as generally known does not suffice to approach the substance of book studies as media studies of the book. I will present the two major arrays of issues – the rearrangement of the value chain on the one hand and media convergence on the other – and after that expose in outlines, how the

VRE (Virtual Research Environment) approach of Johannes Gutenberg University Mainz is set up to support these specific research interests.

The context of the work of the Mainz book studies department is its embedment in a interdisciplinary research unit "Medienkonvergenz" ('Media Convergence'). This research unit brings together humanities and social sciences, disciplines covering media (book studies, journalism studies, film studies, theatre studies) with media law, media pedagogy, media art as well as media management, information science and neurolinguistics, encompassing altogether more than 40 professorships and 5 international research fellows and having a dedicated administrative office at its disposal. This research unit has a role in the application of Johannes Gutenberg University to be one of Germany's 12 top, so called elite universities.

Since in the following we are not talking about texts (possibly with images), but more about media, books, it might be helpful to contour the concept of the book for the purposes of this paper. Books are seen here as artifacts or immaterial phenomena with certain functions: transport of predominantly textual information (possibly complemented by images and other entities from different sign systems) with specific features (length, argumentative complexity, aesthetic "pretension", etc.) easing the burden for the individual or collective memory and inducing publicness, or for certain entertainment, education, information, edification, purposes, respectively. (Bläsi, referring among others to Rautenberg (cf. e.g. RAUTENBERG: 2002, 82-85). When digital forms of books are

mentioned, the focus is mostly on ebooks as the most important digital form. Ebooks are taken as digital 'book-like' (see above for hints with respect to content-related determinants of the concept) files for the use primarily on mobile devices, either dedicated ereaders like Amazon's Kindle or tablet computers like Samsung's Galaxy Tab. For completeness sake, it should be recorded, that other forms of digital books include e.g. Wikipedia, content based services like spell checkers or tax declaration software, but also e.g. (free) HTML offerings like Project Gutenberg.

Research problem area I: The arising value network of the book industry: rearrangement in progress

As the first of the suggested additional object areas for Digital humanities approaches book studies can intrigue fellow humanities researchers, the value chain of the book economy, especially under its current rearrangement to a value network, will be explained in more detail. On the one hand, the traditional players of the century-old book trade – publishers, wholesalers, and booksellers – are widely still in business, partly with modified offerings. The latter can be shown using the example of Thalia, Germany's leading book selling chain with stationary outlets as well as an online bookshop, which now also sells its ereader Oyo and ebooks, or of Amazon (which indeed is not a typical bookseller, but essentially still is a bookseller), which sells its ereader Kindle and ebooks. On the other hand, there are players not traditionally in the book business as well as unexpected joint ventures. E.g. IT and Telecom companies can be mentioned: Apple's iBookstore is online in Germany since early 2010 and Google's ebookstore is announced for the near future. Moreover

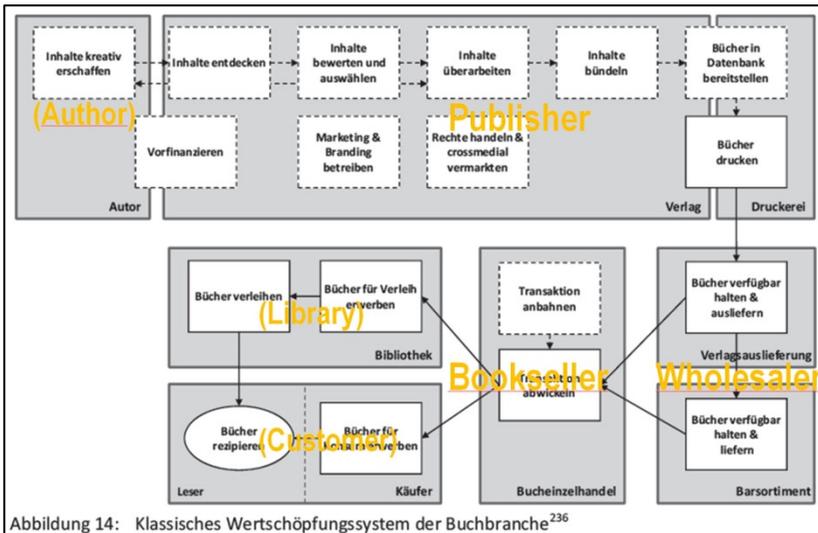
and among many others, there is PagePlace, the "cloud" ebook online shop by telecoms company Deutsche Telekom, Pubbles, the ebook online shop by publishing and book selling conglomerates Bertelsmann and Weltbild/Hugendubel as well as the new intermediary between publishers and ebook shops set up as a cooperation between wholesaler Umbreit, market researcher Media Control, and (originally digital music) distributor Bookwire. Some rearrangements are even more thorough than the ones mentioned so far, potentially truly game-changing: under the label of self publishing, authors turn to their readers without any intermediary at all and with a phenomenon called "semantic publishing", distributed authors record (e.g. „triplified", i.e. in the format suitable for the Semantic Web) statements that get interlinked and form something like a 'fluid' publication, for which there is no clear border between externalisation and publication in a narrower sense any longer.

A few illustrations will give a vivid birds-eye perspective impression on the exciting changes under way. Illustration 1 shows the traditional book value chain according to Janello (cf. JANELLO: 2010), illustration 2 Darnton's (cf. DARNTON: 1982) earlier book communications circuit model – the latter with a strong universality claim¹. The view depicted and postulated, respectively, by these is

¹ [...] [B]ooks generally pass through roughly the same life cycle. It could be described as a communications circuit that runs from the author to the publisher [...], the printer, the shipper, the bookseller, and the reader. Authors are readers themselves. [...] So the circuit runs full circle. [...] The model [...] provides a way of envisaging the entire communication process. With minor adjustments, it should apply to all periods of the [...] book [...]" (DARNTON: 1982, 67).

rebutted by illustration 3, which shows the arising book value network according to, again, Janello as well as by illustration 4, Darnton's communication circle edited by me for the case of Wikipedia² as a sufficient counter example.

Figure 1: Traditional book value chain³.



² Wikipedia is not least seen as an instance of book communication, since it has ultimately driven out of business one of the leading German publishers, Brockhaus – by means of a digital 'product' that for millions of people substitutes Brockhaus' encyclopaedias heading this field for about 200 years.

³ Source: JANELLO: 2010, 58; translations in yellow by C.B.

Figure 2: ('Universal') Book communications circuit⁴.

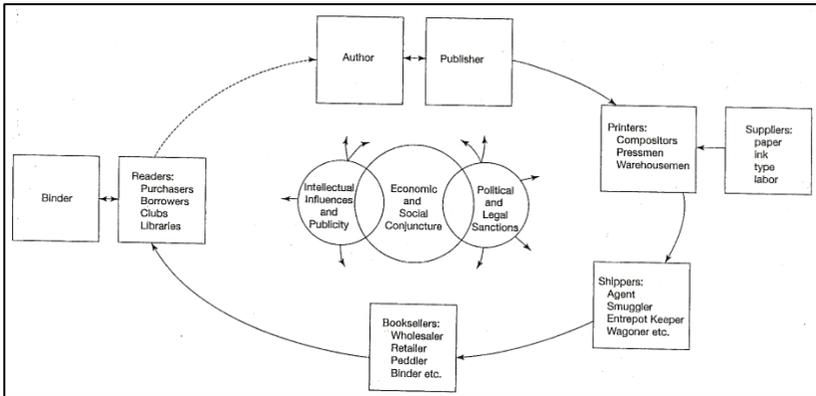
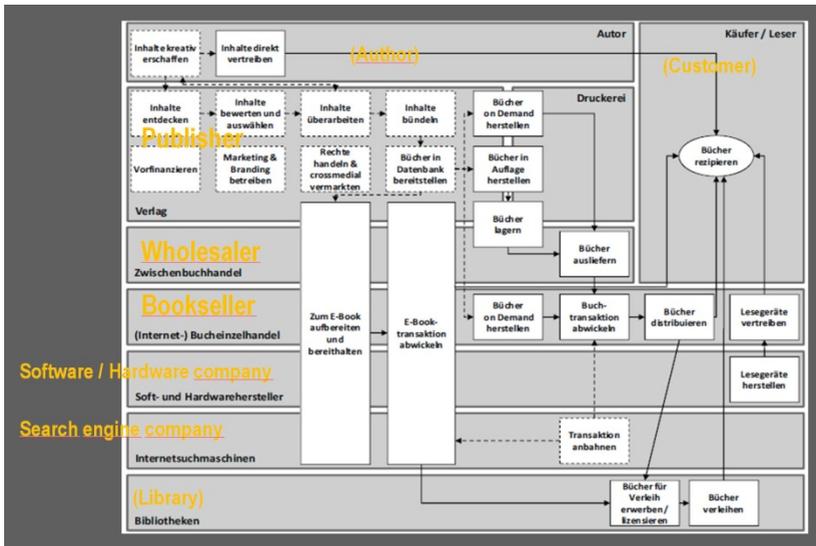


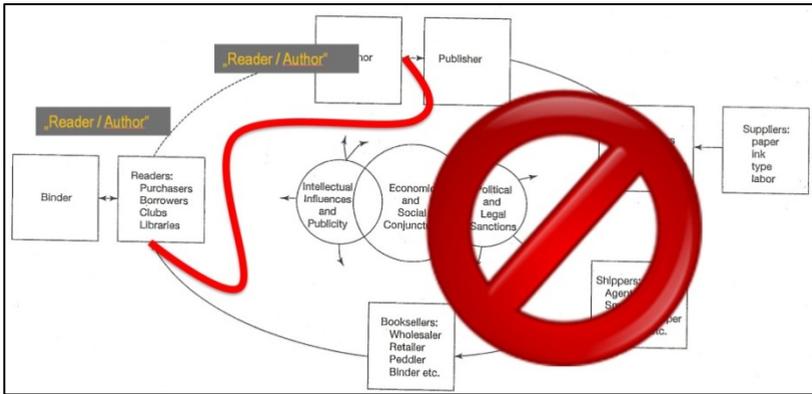
Figure 3: Current book value network⁵.



⁴ Source: DARNTON: 1982, 68.

⁵ Source: JANELLO: 2010, 97; translations in yellow by C.B.

Figure 4: Book communications circuit⁶.



Research problem area II: Media convergence phenomena (tangent to the book)

Media Convergence – also if limited to phenomena tangent to the book – is much more than its effects on the rearrangement of the book value creation; I will present examples that do not have a direct bearing on the value creation below. Then again, not all disintermediation, reintermediation etc. phenomena in the book industry we observe today are directly connected to media convergence, cf. e.g. the typical bypassing of stationary book sellers in the case of ebooks. Nevertheless, media convergence phenomena and phenomena of the rearrangement of the book value creation undoubtedly intersect. In the following, aspects of media convergence in its effects on the medium book and its stakeholders will be explained in more detail.

⁶ Source: DARNTON: 1982, 68, edited in red for the case of Wikipedia by C.B.

Silverstone (cf. SILVERSTONE: 1995) opined pithily about media convergence by entitling a paper "convergence is a dangerous word"; not least based on this bold statement it seems worthwhile to state, what should be meant by media convergence for the train of thought ahead. This shall be done using, to begin with, the focus on technological convergence by e.g. van Dijk and Latzer, a classification of entailed convergence phenomena by Quandt as well as by thoughts by Jenkins on the challenges of media convergence. According to van Dijk and Latzer (cf. VAN DIJK: 2006, LATZER: 1997), technological convergence is the digitally induced convergence of mobile and landline telecommunication, data communication and mass communication. Convergence phenomena entailed by this technological convergence are – according to Quandt (cf. QUANDT: 2008) – organisational, production, content, usage, and audience convergences. Finally, Jenkins (JENKINS: 2010, 41) lists challenges of [media, C.B.] convergence processes: "1) revising audience measurement, 2) regulating media content, 3) redesigning the digital economy, 4) restricting media ownership, 5) rethinking media aesthetics, 6) redefining intellectual property rights, 7) renegotiating relations between producers and consumers, 8) remapping globalization and 9) re-engaging citizens."

Subject to more differentiated statements on the current situation of the book with respect to media convergence I will make below, it is easy to see, that this medium has been affected by media convergence processes only very recently. Digital music and digital films, interconnected by a wide range of convergence ties (cfr. below), have been with us for more than a decade, early digital clips date back much further. Whereas the book has reached a position, in

which it does make sense at all to discuss this matter, only as late as 2007, when the Amazon Kindle ereader had been introduced to the US market. This is not obvious, since there were single phenomena that could be seen as digital / 'electronic' books decades before anyone could think of digitized music or films. Early experiments and widespread related phenomena, especially on-screen computer manuals, based on digitized (long-form "narrative") text date back to pre-PC, i.e. mainframe and workstation times between the 1950s and the 1980s. These attempts had been facilitated by the low bandwidth and computing power requirements of digital texts (in contrast to e.g. music or film). Notwithstanding these observations, the book had become available in formats making it amenable for media convergence, i.e. digital ones, to a larger extent only a few years ago, with the advent of the Kindle.

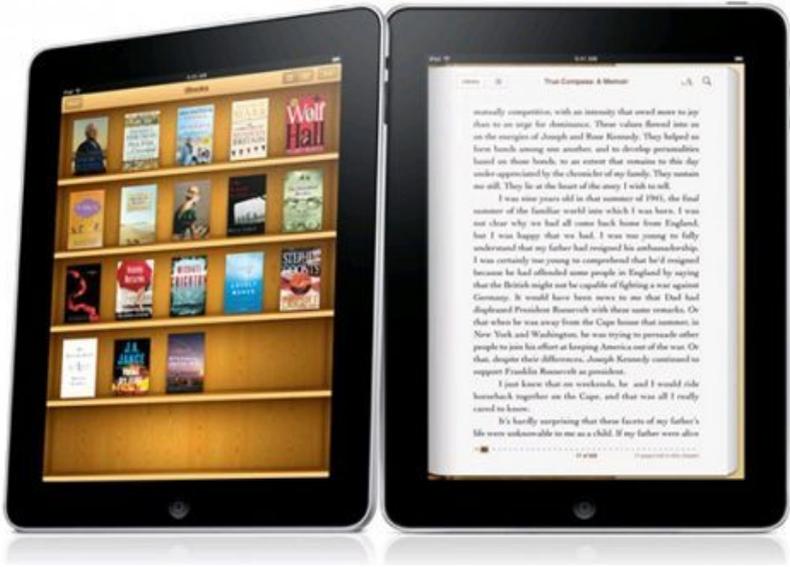
The following is structured on the top level along the types of entailed convergence phenomena according to Quandt; going through the various phenomena, it will be examined in each case, if the book as a medium is affected by the kind of convergence in question⁷. On the basis of a striking example in each case, it will be possible to maintain, that the medium book is affected by media convergence in the respective dimension. For some of the dimensions, challenges in the wake of convergence (according to Jenkins) will be discussed briefly.

⁷ This will be done referring to the paper "The book – civilizing forerunner and media convergence latecomer: phenomena, challenges and a suitable research infrastructure" the author of this paper has given at the "Media Convergence" conference at the Jesuit University of Kraków, Poland, on March, 17th, 2011.

Digitally induced ebooks are – quite trivially, because this is a constitutive part of the concept as it has been introduced above – digital; hence, the technological prerequisite of any (digitally induced, cfr. Latzer) 'higher level' (media) convergence phenomena is fulfilled.

Organisational Convergence. Apple Inc. is originally a hardware manufacturer and a software company. Over the last years, however, it has developed also into a retailer of digital content, not only of music, but – by means of its iBookstore portal – of books, too. This is an instance of organisational convergence: the organizations behind the distribution of different (digital) media (in this case particularly music and books) are the same. Based on Jenkins, the following challenges can be formulated: redesigning the digital economy, remapping globalization – and maybe restricting media ownership. Traditional enterprises in the book trade – in this case especially wholesalers and bookstores – get seriously challenged by a player from outside their industry, in this case even from abroad, entering an industry long thought to be naturally national due to its culture and language dependence. Especially the possible dominance of a player across various media could give rise to demands for ownership restrictions.

Figure 5: Photo of Apple's iBookstore and a 'book' page on an iPad⁸.

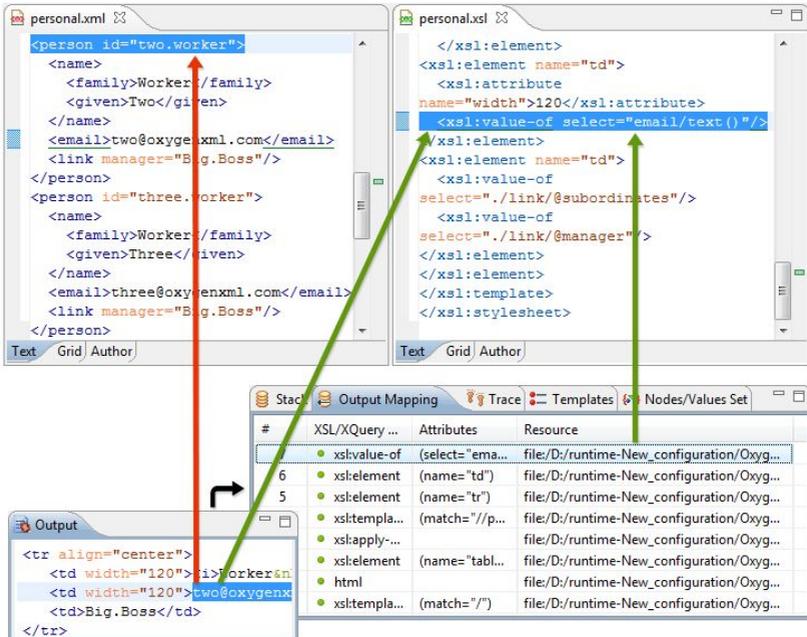


Production Convergence. ebooks can be – and at least ideally and in cross-media environments typically in fact are – "rendered" (i.e. converted) from the same media-neutral XML data repositories in enterprise content management systems as for example the websites of the corresponding media company (this is in an state-of-the-art environment of course also true for the digital [e.g. PDF] files that are handed over to the printer in the case of printed books). The standard cross-platform ebook format ePUB, that in its 3.0 version (finalized in October 2011) is gradually gaining ground, is even itself based on XML / HTML. This is an instance of production

⁸ Source: maclife.de.

convergence: (production) formats and hence asset management approaches and workflows are used across different media (in this case websites and books).

Figure 6: Screenshot of an XML editor screen⁹.

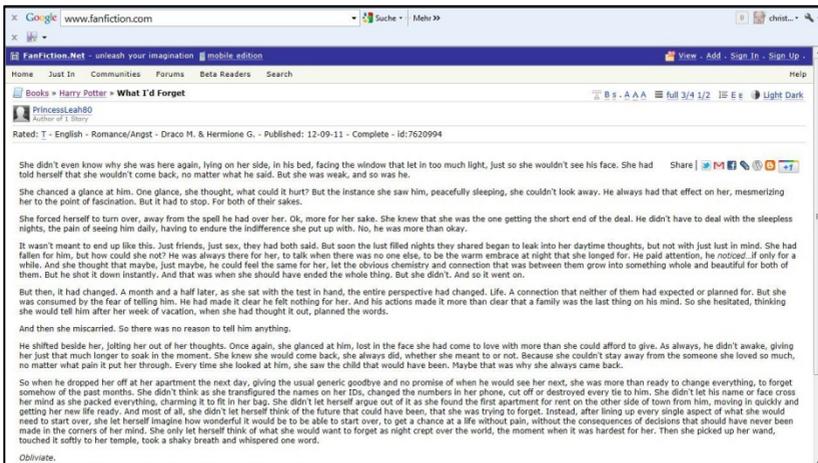


Content Convergence. Fan fiction, especially on corresponding user generated content platforms (in German language e.g. fanfiction.de, in English language e.g. fanfiction.net), makes use of the same story worlds, in particular characters, as the original content it is based on; this base content could have been conveyed as

⁹ Source: oxygenxml.com.

a book – or indeed as an ebook. This is an instance of content convergence: content is developed, modified, refined, expanded, etc. across different media (in this case [social] web and ebooks). Based on Jenkins the following challenges can be formulated: renegotiating relations between producers and consumers – and also redefining intellectual property rights. Fan fiction could be a productive part of a participative story culture – copyright regulations permitting. Moreover, fan fiction can be a means for talent sourcing, facilitating role changes from consumer to (professional) producer.

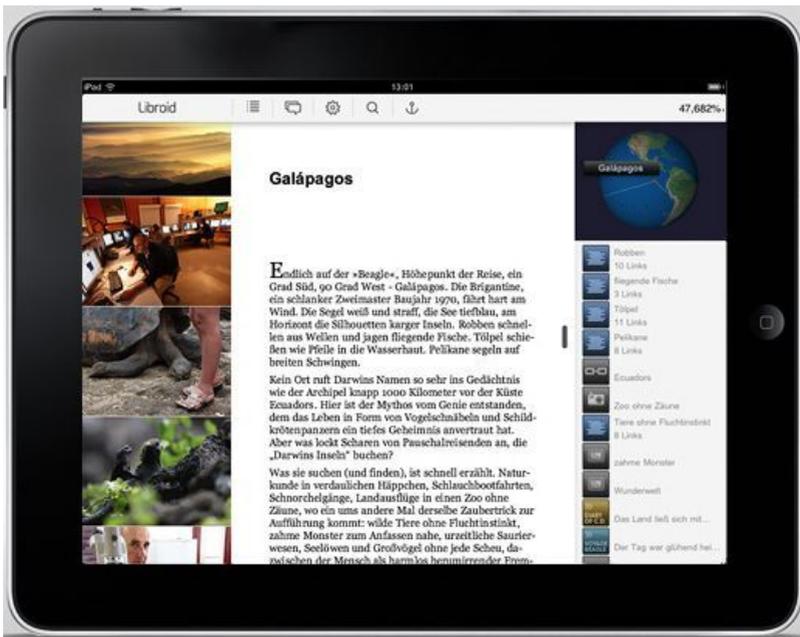
Figure 7: Screenshot of an entry on fanfiction.net.



Various Convergence Phenomena. 'Enhanced' ebooks incorporate content types (e.g. videos, audios, 'interactivities' / simulations), but also e.g. search options, active links and location based services formerly associated with media like cinema, (mobile) web, etc. This is an instance of several entailed convergence phenomena and it has

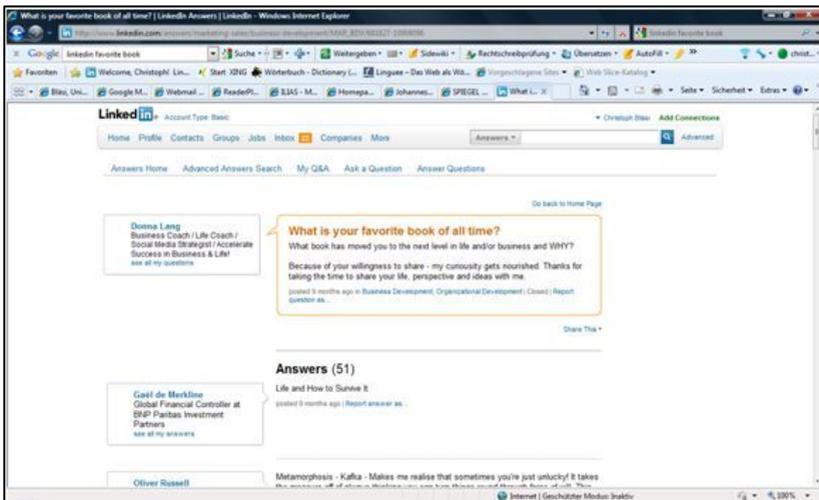
expectedly also a whole set of challenges in store, not least – this again based on Jenkins – rethinking media aesthetics. Convincing enhanced ebooks are about to establish new forms of artistic expression beyond the mere compilation of different media (cf. the transition from the silent movie via the silent movie accompanied by music to the 'talkie' / sound film from the 1920s to the 1930s): new, complex kinds of telling stories using the modes of expression available for 'enhanced' ebooks are evolving.

Figure 8: Photo of a page of the 'enhanced' ebook / "Libroid" "Darwin" by Jürgen Neffe. Source: siteseeing.de.



Usage / Audience Convergence. Favorite books are used in social networks much the same way as favorite music – by the same people and potentially on the same (mobile) devices, quite likely with identical fundamental functions (e.g. featuring oneself or finding soul mates). This is an instance of usage / audience convergence: the same people make the same use of different media (in this case of music and of books).

Figure 9: Screenshot of a respective entry on linkedin.com.



Complications and related terms. The example cases above indicate that 'high-level' media convergence phenomena involving the book can indeed be spotted easily – 'high-level' here in the sense of converging with respect to more than just involving content streams of strings of 0s and 1s, i.e. digital content streams. Not only related to the book, the wider object area is more complicated than this,

however: as entailed phenomena of the technological *con*vergence (based on digitization, see above) also opposed, i.e. *di*vergence, phenomena can be stated, e.g. an audience *di*vergence in socialization processes, for example based on distinct gender differentiation. An example of this would be types of computer games played predominantly by girls as opposed to ones predominantly played by boys. As humanities scholars we have to record, moreover, that phenomena of media convergence in a narrower sense of the concept cannot always clearly be separated from phenomena of (the wider concept of) intermediality, based on which strand of research the relations between two phenomena in instances of two distinct media can be described more thoroughly than just as being a case of content convergence (if they are the latter in the first place – the wider concepts should of course not least be and in fact is applicable also to 'analogue' forms of content !). Among such intermediality (cf. RAJEWSKY: 2002) phenomena are e.g. those of media in contact in a very broad sense, considering e.g. story fragments wandering through epic poems, songs and plays in oral cultures. Also from an economics / management point of view, "content convergence" can be too imprecise a concept – considerations referring to the multiple exploitation of intellectual property in the book / media industry in more detail (using e.g. concepts as versioning or windowing) are often of much more explanatory power than just looking at phenomena as instances of media convergence.

An important contribution to tackling these research problem areas: "The future Mainz VRE infrastructure"

As could be shown, there are relevant research problem areas in the scope of contemporary book studies – a so far largely undisputed humanities discipline – that are data-rich and methodologically demanding, the rearrangement of the book value chain / network and media convergence had been accented among them. Methodologically demanding means, that not least a greater diversity of methods, beyond purely qualitative and hermeneutic approaches, is required. All this is one of the reasons, why the impetus to put a special emphasis on a future virtual research environment (VRE) for the humanities at Johannes Gutenberg University originated in its book studies department. This impetus in direction of such a VRE has been integrated into the information infrastructure strategy of Johannes Gutenberg University, which has made an excellent information infrastructure a focal point of its further development.

It has been dwelled on in several places of this book, that access to digital resources and publications, the storing of research data, as well as the production, the shared processing, the documentation, and the publication of research results and their long-term archiving are the main purposes of virtual research environments. The adherence to standards of various types has been pointed out as a central requirement. VREs can support inter-/transdisciplinary as well as locally distributed research groups and ideally their services should be integrated in national and international Digital humanities

projects and infrastructures – in the German case e.g. in TextGrid¹⁰ or DARIAH-DE¹¹.

Johannes Gutenberg University has made the VRE concept also an integral part of its application to be one of Germany's top 12 universities – first beneficiaries of the environment will be the research units "Historical cultural studies" and "Media Convergence". Independently from the research problem areas outlined above, the book studies department has devised a project, the aim of which it is to recognize, monitor and interculturality compare cultural trends by observing the book title supply in markets over time with the help of a 'data mining' system, using among others Semantic Web and Linked Open Data technologies. In the centre of this system there will be a big data cube compiled from numerous data sources on books in the web from libraries and social reading platforms to industry association databases and online bookshops. This project will be the first instance, for which the VRE support will explicitly expanded to book studies research problems, tackling 'material' aspects of media.

¹⁰ <http://www.textgrid.de>.

¹¹ <http://www.dariah.eu>.

Conclusions

The application of Digital humanities approaches should not be limited to philological research problems in a narrow sense. Book studies as a humanities discipline constitute an excellent case to apply such approaches to media studies / materiality subjects as the relevant and complex phenomena around the book as a medium presented above prove.

These phenomena – themselves of course mediate consequences of digitization – are to be researched inter/transdisciplinary and internationally and include a whole lot of items to be processed digitally (primary files, publications), partly with the help of advanced, 'semantic' data structures and algorithms. All this makes them ideal objects for VRE-supported research.

Johannes Gutenberg University of Mainz has made the development and deployment of a state-of-the-art VRE for the humanities one of the focal points of its further development.

Riferimenti Bibliografici

- Ch. Bläsi, *E-Books und die Stakeholder des Buches*, in *Information. Droge, Ware oder Commons? Wertschöpfungs- und Transformationsprozesse auf den Informationsmärkten. Proceedings des 11. Internationalen Symposiums für Informationswissenschaft (Schriften zur Informationswissenschaft, Bd. 50)*, Boizenburg, Hülsbusch, Rainer Kuhlen, 2009, p. 335-348.
- R. Darnton, *What is the History of Books?*, in «Daedalus», 111, 1982, 3 (*Representations and Realities*), p. 65-83.
- Ch. Janello, *Wertschöpfung im digitalisierten Buchmarkt (Gabler Research: Markt- und Unternehmensentwicklung)*. Wiesbaden, Gabler, 2010.
- H. Jenkins, *Convergence culture: where old and new media collide*, New York, New York University Press, 2010.
- M. Latzer, *Mediamatik – die Konvergenz von Telekommunikation, Computer und Rundfunk*, Wiesbaden, VS, 1997.
- Th. Quandt, *Medieninnovationen und Konvergenz. Formen, Faktoren und Felder des Medienwandels. Habilitationsschrift*, München, Ludwig-Maximilians-Universität, 2008.
- U. Rautenberg (ed.), *Reclams Sachlexikon des Buches*, Stuttgart, Reclam, 2003.
- I. O. Rajewsky, *Intermedialität*, Tübingen, Francke / UTB, 2002.
- R. Silverstone, *Convergence is a dangerous word*, in «Convergence», 1, 1995, p. 11-13.

Automi e linguaggio nell'ecosistema delle reti digitali

Stefano Lariccia (DigiLab Sapienza)

G.Toffoli (Link)

In questo lavoro di rassegna vogliamo presentare alcuni strumenti interessanti nell'ambito del NLP (Natural Language Processing) per esemplificare quali obiettivi ci si può proporre di raggiungere, quali strumenti abbiamo a disposizione, quali raffinamenti è possibile prevedere e in quale arco temporale. Insomma, si potrebbe riassumere, quale futuro ha la futurologia? Accanto alle opportunità che questo panorama offre agli studiosi del linguaggio e delle applicazioni delle tecnologie dell'informazione nel dominio delle scienze umane, emergono anche grandi interrogativi etici: è possibile accettare che il progresso tecnologico avanzi seguendo meccanismi autonomi di avanzamento? È immaginabile proporre di istituire regole di governance per l'ecosistema linguistico concettuale del WorldWideWeb? In occasione del convegno sull'opera di Giuseppe Gigliozzi, pioniere nella applicazione delle scienze del computer all'interpretazione del testo, ci si offre

l'opportunità di analizzare come e quanto siano progredite le prospettive, anche in termini educativi, di questo settore degli studi linguistici nei venti anni trascorsi.

WebScience: la scienza del futuro prossimo, al crocevia tra tecnologia e comunicazione, tra storia, letteratura e futurologia

Il Web è intrinsecamente un organismo composto di "engineering structures" e di "social negotiation structures". Nessuno potrà riprodurre il successo straordinario dell'originale WorldWideWeb, nemmeno le stesse menti che lo immaginarono nel '90, se non si creerà lo stesso "consenso" intorno alla nuova idea.

In *A Framework for Web Science*, che è l'ultimo lavoro stampato di Tim Berners-Lee, inventore del WorldWideWeb nel 1989 e del Web Semantico nel 1999, il più costoso, sintetico e produttivo testo stampato¹, almeno a mia memoria, alcune indicazioni degli autori sono estremamente chiare riguardo a questo argomento:

Web Science is deliberately ambiguous phrase: physical science is an analytic discipline that aims to find laws that generate or explain observed phenomena; computer science is

¹ Il libro è composto di 134 pagine ed ha un prezzo di copertina di 90 euro + costi di spedizione (totale 110 euro); in base al lavoro indicato in questo libro (ma ovviamente la mia è una provocazione, il risultato che cito nel seguito è molto più dovuto all'incredibile lavoro organizzativo svolto dal gruppo di autori nell'ambito del W3C per circa 15 anni) il governo inglese ha stanziato nell'aprile 2010 circa 45 milioni di pound per dare il via ad un progetto di ricerca che viene definito da diverse fonti come "una via britannica al Semantic Web".

predominantly (though not exclusively) synthetic, in that formalisms and algorithms are created in order to support a particular desired behavior. Web Science has to be a merging of these two paradigms; the Web needs to be *studied and understood*, and it needs to be *engineered* (p.2)

Dunque una scienza che in modo del tutto nuovo e interdisciplinare dovrebbe *apprendere* a coniugare, in una *nuova agenda della ricerca*, le proficue interazioni che si devono attivare ai vari livelli per facilitare le proprietà desiderabili e disincentivare le proprietà non desiderabili.

Applicazioni già note e applicazioni concepibili a breve:

- Glattfelder e Battiston²: La fotografia dei "poteri forti": analisi delle interconnessioni tra le 43.000 principali organizzazioni finanziarie del mondo.
- Le previsioni criminologiche fatte dai servizi segreti "interrogando" i materiali apparsi sul web.
- Le previsioni econometriche basate sull'analisi dei feed del web.
- Il WAC: Web as a corpus; il corpus linguistico-letterario costituito dall'insieme del web utilizzato come campione statistico in riferimento ad una lingua naturale.
- Sistemi di accreditamento aperto e interoperativo.
- Da Wikipedia a dBpedia: una gigantesca base di dati a disposizione del progetto di un Thesaurus integrato multilingue.

² S. Vitali, J. B. Glattfelder, S. Battiston, "The network of global corporate control", *PloS one* 6 (10), e25995.

- Sistemi "aperti" di tracking sul web in generale per determinare il valore di un prodotto culturale e riportare l'impact factor (Hirsh Number) di un autore sul dibattito in un determinato settore scientifico.
- Sistemi aperti per monitorare le competenze acquisite attraverso l'esplorazione di risorse formative sul web.
- Sistemi aperti per assegnare una certificazione "interoperabile" delle competenze acquisite.

Strumenti per la estrazione di senso dai testi pubblicati sul Web. Il corso di Computational Linguistics alla Sapienza per il CdL di Editoria, di Mediazione linguistica e per la Scuola Speciale Archivistici e Bibliotecari.

Gli autori di questo lavoro sono responsabili della organizzazione e dei materiali del corso che verrà tenuto da S.Lariccia presso questa università nel secondo semestre il corso "Informatica e Linguistica per Scienze Archivistiche e Bibliotecarie". Il corso è articolato in un laboratorio di Computational Linguistics e in un corso monografico che affronta lo sviluppo di Internet nei suoi primi 40 anni e del WorldWideWeb nei suoi 20 anni di vita. Questi artefatti hanno trasformato – in modo mai avvenuto prima d'ora nella storia dell'uomo, le pratiche, le tecnologie e persino gli obiettivi fondamentali di molti dei mestieri tradizionali cui un giovane, portatore di cultura non "specialistica" era normalmente indirizzato. Il mestiere di insegnanti, educatori, giornalisti, archivisti e bibliotecari è stato radicalmente trasformato.

Questo modulo avanzato intende fornire una visione d'insieme storicizzata della evoluzione degli strumenti tecnologici al servizio delle discipline tradizionalmente definite "umanistiche". In particolar modo si vuole rianalizzare criticamente il ruolo che tali discipline "umanistiche" hanno tradizionalmente interpretato e continuano ad interpretare nel coordinare le necessità di documentazione, classificazione, archiviazione, validazione dell'informazione nei circuiti tradizionali e innovativi di creazione e diffusione delle conoscenze. L'incremento di aspettative verso una sempre più ampia accessibilità delle nuove conoscenze; le aspettative verso una sempre maggiore democratizzazione dell'accesso alle conoscenze; le aspettative per una sempre maggiore efficacia nei metodi di validazione e controllo delle fonti di informazione: questi sono alcuni tra i più evidenti risultati della rivoluzione del Web nella gestione sociale delle conoscenze. Fino a che punto queste originarie aspettative sono state realizzate nei primi anni di diffusione globale del WorldWideWeb? Quale è la frontiera attuale su cui si lavora oggi e quali politiche, quali strategie dovrebbero essere implementate per dare vita ad ulteriori successi in questa direzione? Il governo delle tecnologie al servizio della conservazione, identificazione ed elaborazione della cultura immateriale, sulla quale si basa l'elaborazione storica e politico-progettuale è divenuto una condizione fondamentale per assicurare equilibrio al progresso della scienza e della stessa condizione umana. In questo ambito sono divenute fondamentali nuove tecnologie legate alla elaborazione dei contenuti elettronici: elaborazione del linguaggio naturale; estrazione di contenuti semantici dalle collezioni di documenti

digitali; strumenti di classificazione su basi semantiche; utilizzazione di ontologie al servizio degli strumenti di classificazione.

Nel corso verranno anche presentati, attraverso una opportuna bibliografia, casi di documentazione ed elaborazione dell'informazione "diffusa": in particolare verranno presentati casi di "crowdsourcing", o di "elaborazione sociale" come alcuni casi inseriti nell'iniziativa del governo federale americano per la diffusione dei dati statistici "Open Data Initiative"³. Questi strumenti verranno presentati a partire da una utilizzazione pratica e in un ambiente sperimentale e verranno impartite le conoscenze tecniche per accedere a questo settore in via di sviluppo dell'informatica e della mediologia.

Un manifesto per l'apertura delle risorse di Natural Language Processing nella comunità italiana dei ricercatori in area umanistica

Con COACH, un set di strumenti accessibili trough-the-web, un workbench, per l'analisi dei testi e l'estrazione di termini da corpus linguistici pre-esistenti in italiano, inglese e spagnolo intendiamo presentare un "manifesto" per la costituzione di una comunità di ricerca attorno ad uno strumento, un workbench appunto, che ci consenta di aggregare le migliori risorse disponibili attorno ad alcune linee guida della ricerca in area umanistica. Il presente lavoro

³ The Open Data Initiative è una iniziativa ospitata sui server del Google App Cloud Engine che invita a condividere e a rendere accessibili in modalità Web 2.0 i dati di rilevanza pubblica.

è proposto come un work-in-progress, consistente nell'adattamento, nell'estensione e nell'integrazione attorno all'ambiente Plone di strumenti per l'elaborazione di testi (NLP), per il tagging, il crunching e sullo sfondo l'inferenza semantica a partire corpora pre-elaborati contro cui confrontare testi singoli, e raccolte ulteriori di testi. Il focus del lavoro sarà per ora limitato (per quanto riguarda la sperimentazione iniziale) sulla lingua italiana. Il primo obiettivo è infatti quello di disporre anche per l'italiano di una suite di tool che consentano di effettuare una migliore indicizzazione full-text dei documenti e di estrarre da essi i "termini" rappresentativi candidati ad essere usati per il "tagging" dei contenuti stessi (nel contesto di blog, digital library, ecc.). In seguito vorremmo supportare costruzione semi-automatica di glossari e ontologie, web mining, ecc. Il secondo obiettivo è quello di far tesoro delle risorse generate – in maniera più o meno consapevole nel workbench comune – dalle attività autonome di ciascuno degli stakeholders, da ciascuno dei membri della comunità di utenti. Il terzo obiettivo è quello di costituire corpus specializzati nell'esame di testi non contemporanei, introducendo variabili e modelli che consentano l'identificazione su una scala diacronica dei testi sottoposti a indagine.

Il panorama internazionale e la situazione italiana

Alcuni tool di questo tipo sono disponibili per la lingua inglese, sia sotto forma di librerie/package per Python o Plone, sia sotto forma di servizi su web; ma sono praticamente inesistenti (o meglio, inaccessibili pubblicamente) per l'italiano.

I principali componenti che intendiamo integrare in Plone sono:

- NLTK (Natural Language ToolKit), una libreria Python e un insieme di risorse linguistiche, soprattutto corpora, risultato della più interessante iniziativa open-source nel campo della linguistica computazionale; si tratta di divulgazione ad alto livello;
- un paio di risorse linguistiche di libero dominio e di ottima qualità disponibili per l'italiano: costituiscono un'eccezione in un panorama desolante, tenuto conto anche degli ingenti finanziamenti pubblici italiani ed europei che sono stati dedicati a questo settore nel periodo 1990-2010.

Scelta di linguaggio artificiale e di ambiente applicativo

A parte PERL, che è un linguaggio di scripting specialistico e dalla sintassi criptica, crediamo che Python sia il linguaggio di programmazione general-purpose che meglio supporta la sperimentazione nel campo dell'analisi dei testi.

Plone di per sé non aggiunge un contributo essenziale, ma

- un sito Plone potrà beneficiare del package in corso di sviluppo;
- la piattaforma Zope/Plone è comunque un'ottima base di partenza per qualsiasi applicazione che deve essere accessibile su web; e le funzionalità di content management e di user management di Plone faciliteranno la gestione di risorse linguistiche personalizzate (es: basi di documenti e file di parametri) nell'ottica di installazioni che forniscano un servizio aperto.

Definizione della scelta di base: Python NLTK + Plone per generare una comunità di utenti per la Linguistica computazionale

Nell'implementare COACH, l'ambiente collaborativo di NLP, abbiamo definito alcuni obiettivi *funzionali* che ci sembrano abbastanza innovativi e che allo stesso tempo costituiscono un primo step realistico per "innestare" COACH nell'utilizzazione pratica di un numero consistente di apprendisti e di ricercatori. Gli obiettivi elencati in ordine non gerarchico di importanza, sono in prima approssimazione i seguenti:

1. Effetto Wikinomics: Integrare tramite un CMS – a partire dal progetto NLTK – la capacità di feedback intelligente sugli algoritmi di apprendimento basati sul confronto con un corpus standard - ad una comunità di studenti in linguistica computazionale, in modo da usufruire del supporto di un numero di operatori umani superiore di 1 o più ordini di grandezza rispetto agli attuali ristretti gruppi di ricerca (da 100 a 10.000, da 100 a 100.000).

2. Rendere accessibile a linguisti, filologi, studiosi del testo, storici della lingua italiana e storici uno strumento che sia in grado di fornire un servizio valutabile per un ampio numero di ricercatori, rendendo così appetibile il suo utilizzo e creando in questo modo un feedback positivo sia per il perfezionamento di un corpus e delle regole di campio-namento del corpus, sia per il perfezionamento degli algoritmi di analisi ed estrazione termini.
3. Mettere a test la necessità di potenza di calcolo, sperimentando diverse soluzioni per rendere disponibili potenze di elaborazione crescenti on-demand, rendendo possibile la profilazione, la rendicontazione ed il pagamento del computing power richiesto da ciascuna ricerca.
4. Identificazione di uno standard identificativo unico per i termini di una lingua (sull'esempio del DOI? dei riferimenti bibliografici? ecc.) che renda possibile la referenziazione dinamica, basata su una autorità formale, di un termine, di un concetto, di un significato in uno spazio linguistico definito da una sorgente – autoritativa (Webservices DNS?); provvedere ad un meccanismo di classificazione diacronica di tale lista.

Obbiettivi applicativi innovativi di COACH all'interno del progetto Sapienza "Emozioni"

Un diverso insieme di obbiettivi funzionali che abbiamo tenuto in mente nell'implementare COACH, possono essere definiti piuttosto obbiettivi *applicativi*. Questi, al contrario degli obbiettivi *funzionali*, sono il primo banco di prova per testare la possibilità di rendere utilizzabile COACH in tempi sufficientemente rapidi perché avvenga il bootstrap della comunità di utenti necessario per vitalizzare il progetto. Anche in questo caso abbiamo individuato obbiettivi che ci sembravano sufficientemente realistici (ciò che linguisti e filologi si attendono da uno strumento reale ed esistente) e allo stesso tempo potenzialmente innovativi. Gli obbiettivi elencati in ordine non gerarchico di importanza, sono in prima approssimazione i seguenti:

1. Essere in grado di identificare le probabilità di occorrenza e co-occorrenza di termini dati in un contesto specifico dato, a fronte di analoghi valori generali dell'Italiano (classificazione degli ambiti della langue); significa avere meccanismi automatici di pre-identificazione di un numero *limitato* di contesti linguistici; crono-localizzazione.
2. Mettere un ampio numero di utenti proattivi in grado di sperimentare algoritmi per l'associazione di stili autorali con testi dall'autore non identificato e di fornire feedback per il perfezionamento degli algoritmi stessi; ipso-identificazione.
3. Mettere un ampio numero di utenti proattivi in grado di sperimentare algoritmi per l'associazione di stili regionali e di altre comunità locali e di fornire feedback per il perfezionamento degli algoritmi stessi; geo-localizzazione.

4. Preparare la piattaforma *collettiva* per il successivo ampliamento funzionale che estenda le funzionalità di analisi al dominio della semantica e delle ontologie, mediante la predisposizione di un impianto architeturale software aperto, modulare, documentato e facilmente documentabile attraverso l'attività dello stesso utente.

Possibili casi d'uso concreti nel progetto "Il linguaggio delle emozioni" in altri ambiti

Come primo caso d'uso, all'interno del progetto PRIN "Il Linguaggio delle emozioni" ci proponiamo di costituire un ambiente di "sperimentazione linguistica" (Linguistic Workbench) con il quale mettere alla prova i seguenti procedimenti e i seguenti framework di riferimento per una analisi semantica di corpora letterari:

- individuazione dei radicali caratterizzanti di uso più frequente del corpus SapienzaEmozioni
- descrizione del loro inserimento all'interno dell'ontologia Wordnet per creare una topografia specifica, verificarne il coinvolgimento di altri campi, e la discendenza relativa (facendo emergere proprietà specifiche come funzioni metaforiche e altri usi)
- fare emergere dal contrasto tra corpus testuale (corpus oggetto di indagine) e corpus di riferimento le proprietà specifiche di ciascun dominio indagabile (autore, epoca, lingue, genere, linee stilistiche, "scuole", etc.
- esser pronti a individuare tramite il feedback nuove strategie di studio e nuovi possibili domini.

Possibili casi d'uso di COACH in altri ambiti di ricerca

- Dato un dominio o una sottorete, si chiede al software di analizzare le pagine pubblicate sotto le URI appartenenti al dominio specificato estraendone sintagmi, termini e coppie di termini statisticamente rilevanti;
- Il "corpus delle emozioni" un' analisi attraverso NLTK su un corpus preparato per il gruppo di lavoro;
- Un corpus delle leggi costituzionali di alcuni paesi europei;
- Il corpus degli scritti di un eminente politico dell'Italia del dopoguerra (progetto in fase di realizzazione e perciò per ora, riservato).

Il WAC: Web as a Corpus. Il Laboratorio Virtuale Digilab Virtual Desk

Il laboratorio che utilizzeremo viene definito come Digilab Virtual Desk e sarà realizzato secondo le moderne pratiche che prendono il nome di "Cloud Computing"⁴ per raggiungere i seguenti obiettivi:

- Virtualizzare al massimo le risorse utilizzate, vedendole come modelli astratti di computer per:

⁴ Amazon offre sul mercato EC², ovvero Elastic Cloud Computing che ha tutte le caratteristiche desiderabili e una documentazione ricca e accessibile. Altri provider anche italiani verranno confrontati con quel modello.

- 1) minimizzare i costi riducendoli quanto più possibile all'uso effettivo di banda e di capacità di memoria;
 - 2) minimizzare il tempo ed il costo della messa in opera e del "congelamento" dell'ambiente di calcolo quando non è più necessaria la sua utilizzazione di massa;
 - 3) mantenere la possibilità di aumentare in tempo reale, le risorse di calcolo mano a mano che la domanda di informazione e calcolo aumenti da parte dell'utenza effettivamente interessata (computing power on demand);
 - 4) mantenere, per scopi didattici la possibilità di moltiplicare il numero dei server durante i periodi di formazione replicando secondo i bisogni una immagine contenente tutti gli ingredienti necessari.
- Progettare formazione per l'utilizzo delle risorse virtuali;
 - Progettare un ambiente che consenta un passaggio indolore dalla esercitazione didattica allo stage, sino alla formazione on-place, resa possibile con accordi e convenzioni con aziende interessate ad integrare giovani lavoratori formati su tecnologie di avanguardia.

DVD, Digilab Virtual Desk è pensato per rispondere a questi obiettivi. È un ambiente online realizzato per consentire stage on-line di editing digitale per studenti e laureandi, consentendo a costoro di creare propri ambienti di sperimentazione omologhi all'ambiente ufficiale, consentendo ad imprese interessate a fruire di contenuti digitali su temi specifici di "commissionare" il lavoro di scouting in rete, di redazione e di classificazione dei contenuti. Sui server virtuali utilizzeremo prevalentemente sistemi operativi Linux

Ubuntu, linguaggio di sviluppo Python, Framework applicativo Zope, e CMS Plone. L'ambiente NLTK completerà la suite di prodotti integrati e sarà resa disponibile per consentire l'uso e la sperimentazione di strumenti di elaborazione linguistica, tanto in fase di formazione, che in fase di auto-formazione e di stage virtuale "presso" le aziende.

Conclusioni sul corso di Linguistica e Computer

Scopo del corso è quello di sviluppare le capacità di orientamento e di autoapprendimento in un settore in rapida evoluzione come quello delle tecnologie per l'elaborazione dei contenuti digitali. Un ulteriore obiettivo è quello di sensibilizzare alla miscela di opportunità e rischi che ogni scelta in questo dominio di conoscenze comporta per il bene comune.

Applicazioni possibili degli strumenti di estrazione automatica di significato e della elaborazione automatica dei contenuti del Web: quali obiettivi nei prossimi 5 anni?

Forse può essere visto come un esercizio eccessivamente ambizioso quello di prevedere quali sviluppi possiamo aspettarci in settori cos' dinamici come quello di cui stiamo parlando, nei prossimi cinque anni. Eppure questo è ciò che servirebbe fare per plasmare la didattica a misura delle prevedibili opportunità dei nostri allievi che verosimilmente cercheranno un impiego da qui a 3-4-5 anni.

Nei prossimi 5-6 anni, applicando il tasso di crescita che abbiamo visto all'opera nello scorso decennio, e introducendo il

fattore definito da Kurzweil⁵ come la legge dei ritorni accelerati, è lecito attendersi una proliferazione di applicazioni automatiche nel settore delle previsioni econometriche (a favore degli operatori dei mercati, degli operatori finanziari, ma anche a favore di produttori ed erogatori di servizi nel settore privato. È probabile altresì che la capacità dei servizi pubblici di interagire operativamente con questi operatori continui a diminuire. Se escludiamo la possibilità di rivolte "luddiste" contro tali meccanismi, improbabili in quanto la bandiera della "openness" di Internet e del web consente a questi "agenti opportunisti" di operare nell'alveo di un tessuto che è oramai considerato "benefico" per la società, se escludiamo dunque svolte radicali, le persone saranno sempre più esposte alle conseguenze dell'operato di questi sistemi automatici e sempre maggiore sarà la necessità, per ciascun individuo, di tutelarsi con i propri scarsi mezzi contro un potere scarsamente visibile, scarsamente definibile, eppure insidioso per la libertà e per i diritti delle persone umane.

Educazione al riconoscimento della presenza di automi, di intelligenza automatica; educazione all'uso a propria tutela e a proprio vantaggio dell'intelligenza umana; educazione alla tutela dei propri diritti nel campo del diritto attivo e passivo all'informazione, sia nel senso della utilizzazione ai propri fini di questi meccanismi, mano a mano che saranno resi accessibili a persone con un livello di cultura tecnologica medio e medio-alto; capacità di associarsi in gruppi per meglio fronteggiare l'ambiente delle reti popolato da automi intelligenti. Questi sono alcuni degli obiettivi che riteniamo doveroso assumere e fare propri da parte dei ricercatori e dei

⁵ R. Kurzweil, *La singolarità è vicina*, Milano, Apogeo, 2008.

maestri impegnati nel fornire supporto alle nuove generazioni per affrontare l'orizzonte tecnologico e culturale del 2020.

Un'analisi dei possibili sviluppi dell'ecosistema linguistico alla luce delle teorie di Kurzweil del ritorno accelerato

La convergente disponibilità di:

- risorse di calcolo virtualizzate (Cloud computing, Grid Computing);
- CMS in grado di interoperare grazie a modelli universali di contenuti, indipendenti dalla piattaforma;
- standard interoperabili di "crawler" e di "web-robot" capaci di interrogare il web indipendentemente dalle piattaforme di supporto per i contenuti stessi

renderà presumibilmente disponibile un ambiente nel quale utenti con competenze specializzate (ma non "esoteriche") saranno in grado di occuparsi di compiti complessi e di servizi avanzati da vendere con le proprie agenzie indipendenti o attraverso la mediazione di società di capitale; ad esempio:

- Online Placement Globale di prodotti, servizi, conoscenze, brevetti;
- Indagini di mercato;
- Studi socio economici con ricorsione periodica programmata;
 - Analisi narrativa applicata al mercato multimediale; ricerca del plot più gradito al pubblico di una determinata rete televisiva o di un determinato canale web.

Opportunità di potenziamento delle capacità individuali; rischi e motivi di allarme sociale

In conclusione sembra inevitabile, pur accentuando i limiti delle nostre possibilità di preveggenza, interrogarsi anche sulle opportunità offerte dalla diffusione "popolare" di strumenti automatici di harvesting, racking, publishing e sui rischi connessi alla diffusione futura a livello di massa di questa particolare tecnologia. Più in generale queste osservazioni possono essere estese a molti altri casi che possono essere classificati come il fenomeno possibile della diffusione di massa, priva di controlli, di automi cognitivi. Un caso fra tanti ha conquistato le prime pagine dei giornali negli ultimi anni ed è quello degli algoritmi di trading automatico. La recente crisi iniziata nel 2008 ha fatto parlare di un fenomeno che è iniziato per la verità a diffondersi a livello di "massa" già da oltre un decennio. La possibilità, ampliata dalla diffusione ubiquitaria del web, di lanciare operazioni automatiche di trading di titoli sulla base di algoritmi che interpretano l'andamento del prezzo e prendono decisioni non vigilate da umani, sembra aver avuto un qualche ruolo nella radicalizzazione della crisi del 2008. Non è chiaro quanto l'impatto dell'uso di questi strumenti sia stato solo superficiale o quanto possa aver generato meccanismi di turbativa ad uno strato più profondo. Certamente la speciale sensibilità del campo in cui questi automi operano, con la possibilità di distruggere posti di lavoro, e con essi intere nicchie di esistenza umana, indicano un rischio insopportabile e la necessità di vigilare e prevenire.

Dunque le opportunità di crescita, di progresso, vogliamo usare questo controverso concetto? derivanti dall'innalzamento del livello di automazione di alcuni servizi potrebbero non essere tali da

giustificare il rischio connesso con la perdita di governabilità, specie in taluni campi, delle azioni che avrebbero dirette ripercussioni su persone, su gruppi sociali, su intere società.

Dunque, se si rileva necessario monitorare l'evoluzione dell'ecosistema Web per prevenire eccessi derivanti dalla diffusione di automi e meccanismi autodiretti capaci di trasformare l'ambiente senza un controllo umano diretto, sarebbe forse il caso di creare organismi di controllo, magari basati proprio sulle più sofisticate tecniche di automazione dell'analisi e dell'interpretazione. Insomma sarebbe necessario creare un corpo di vigilanza basato su automi nell'interesse della sovranità dell'essere umano. Ma è necessario soprattutto, come sempre per la tecnologia, a nostro avviso contemporaneamente diffondere conoscenze e competenze su queste tecnologie, in tutti i settori della educazione superiore, assieme alla consapevolezza etica del loro potenziale benefico e malefico.

Riferimenti Bibliografici

AA.VV. Open Data Initiative è un sito Web 2.0 accessibile online tramite la seguente URL: www.opendatainitiative.org

Cass R. Sunstein, Farrar, Straus and Giroux (2009) *On Rumors: How Falsehoods Spread, Why We Believe Them, What Can Be Done - how falsehoods spread, why we believe them, what can be done*

Downey A. B. *How to Think Like a Computer Scientist*, Cambridge University Press

Fuksas, A. 2002, *Etimologia e geografia nella lirica dei trovatori*, Roma.

Kurzweil, R., *The Singularity Is Near: When Humans Transcend Biology*. September. 2006: Penguin (Non-Classics). (trad. Italiana: *La singolarità è vicina*, Apogeo 2008).

Pecheux, M. *Automatic Discourse Analysis (Utrecht Studies in Language and Communication, No 5)*, Rodopi Bv Editions.

Simonetti, N. Filosofia della mente e neuroscienze: il caso dei neuroni specchio, giovedì, novembre 19, 2009, [Relazione SFI, Novara, 28.10.2009](#).

Siorpaes' K. ·Simperl, E. 2010; *Human Intelligence in the Process of Semantic Content Creation* Paper Published online: 2 December 2009 © Springer Science + Business Media, LLC 2009.

Vitali S., Glattfelder J.B., Battiston S. *The network of global corporate control*, PloS one 6 (10), e25995.

