

Antichistica 14
Filologia e letteratura 3

Strumenti digitali e collaborativi per le Scienze dell'Antichità

a cura di
Paolo Mastandrea



Edizioni
Ca' Foscari

Risorse digitali per le Scienze dell'Antichità

Antichistica
Filologia e letteratura

Collana diretta da
Lucio Milano

14 | 3



Edizioni
Ca' Foscari

Antichistica

Filologia e letteratura

Direttore scientifico

Lucio Milano (Università Ca' Foscari Venezia, Italia)

Comitato scientifico

Claudia Antonetti (Università Ca' Foscari Venezia, Italia)

Filippo Maria Carinci (Università Ca' Foscari Venezia, Italia)

Ettore Cingano (Università Ca' Foscari Venezia, Italia)

Joy Connolly (New York University, USA)

Andrea Giardina (Scuola Normale Superiore, Pisa, Italia)

Marc van de Mieroop (Columbia University in the City of New York, USA)

Elena Rova (Università Ca' Foscari Venezia, Italia)

Fausto Zevi (Sapienza Università di Roma, Italia)

Direzione e redazione

Dipartimento di Studi Umanistici

Università Ca' Foscari Venezia

Palazzo Malcanton Marcorà

Dorsoduro 3484/D,

30123 Venezia

Strumenti digitali e collaborativi per le Scienze dell'Antichità

a cura di
Paolo Mastandrea

Venezia
Edizioni Ca' Foscari - Digital Publishing
2017

Strumenti digitali e collaborativi per le Scienze dell'Antichità
Paolo Mastandrea (a cura di)

© 2017 Paolo Mastandrea per il testo

© 2017 Edizioni Ca' Foscari - Digital Publishing per la presente edizione



Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale
This work is licensed under a Creative Commons Attribution 4.0 International License

Qualunque parte di questa pubblicazione può essere riprodotta, memorizzata in un sistema di recupero dati o trasmessa in qualsiasi forma o con qualsiasi mezzo, elettronico o meccanico, senza autorizzazione, a condizione che se ne citi la fonte.

Any part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without permission provided that the source is fully credited.

Edizioni Ca' Foscari - Digital Publishing
Università Ca' Foscari Venezia
Dorsoduro 3246, 30123 Venezia
<http://edizionicafoscarì.unive.it/> | ecf@unive.it

1a edizione novembre 2017
ISBN 978-88-6969-182-9 [ebook]
ISBN 978-88-6969-183-6 [print]

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari - Digital Publishing: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della collana. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.

Scientific certification of the works published by Edizioni Ca' Foscari - Digital Publishing: all essays published in this volume have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Scientific Committee of the series. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

URL <http://edizionicafoscarì.unive.it/it/edizioni/libri/978-88-6969-183-6/>
DOI 10.14277/978-88-6969-182-9



Università
Ca' Foscari
Venezia

**Dipartimento
di Studi Umanistici**



Centro interuniversitario per
l'Informatica Romanistica



Convegno di studio

Risorse digitali e strumenti collaborativi per le Scienze dell'Antichità

Venezia, 2-3 ottobre 2014

Università Ca' Foscari – Aula Mario Baratto

Dorsoduro, 3246

Strumenti digitali e collaborativi per le Scienze dell'Antichità

a cura di Paolo Mastandrea

Sommario

Presentazione

Paolo Mastandrea 11

Considerazioni introduttive

Nicola Palazzolo 15

PARTE 1. COSTRUZIONE E GESTIONE DI CORPORA E PIATTAFORME DIGITALI

Standard e interoperabilità degli archivi digitali archeologici: il progetto ARIADNE

Franco Niccolucci 25

BIA-Net: l'interoperabilità delle fonti giuridiche antiche attraverso il Semantic Web

Daria Spampinato 35

Il progetto EAGLE

Europeana network of Ancient Greek and Latin Epigraphy
e le sue molteplici sfide

Silvia Orlandi 49

Alcune note di ecdotica digitale nella prospettiva della filologia collaborativa e cooperativa

Federico Boschetti 61

Le sfide di una biblioteca digitale del latino tardoantico

Struttura, canone e questioni aperte di codifica e visualizzazione

Alice Borgna, Simona Musso 69

Epigrafia digitale o digitalizzazione delle epigrafi?

Flavia De Rubeis 95

PARTE 2. STRUMENTI DI RICERCA SEMANTICA E PER LA CRITICA TESTUALE

- Dal *Thesaurus* di BIA alle ontologie di BIA-Net
Prospettive di sviluppo e integrazione con altre basi di conoscenze**
Marco Giunti 103
- WordNets per lingue classiche**
Riccardo del Gratta 117
- Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'Information Retrieval**
Stefano Minozzi 123
- Didattica e risorse digitali
Appunti per un percorso sul tema 'tempo'**
Giancarlo Scarpa 135
- Perceptual, Grammatical and Hermeneutical Dimensions of Digitallity**
Giorgio Buccellati 159

PARTE 3. PROFILI ORGANIZZATIVI E GIURIDICI

- La protezione dei dati negli archivi digitali**
Ignazio Zangara 173
- The Tesserae Project
Detecting Intertextuality of Meaning and Sound**
Neil Coffee, Christopher Forstall, James Gawley 189

PARTE 4. UNA TAVOLA ROTONDA: VERSO LA BIBLIOTECA DIGITALE ANTICHIstica

- Introduzione**
Anna Maria Tammaro 195
- Biblioteche digitali e umanisti informatici: nuove partnership?**
Anna Maria Tammaro 197
- Oltre le biblioteche digitali**
Maurizio Lana 205

Biblioteche digitali e specialisti del mondo antico	
Collaborazioni nell'ambito dell'indicizzazione e della ricerca semantica	
Anna Lucarelli	219
Alcuni problemi teorici e pratici posti dalle <i>digital libraries</i>	
Raul Mordenti	231
Nota conclusiva	
Raffaele Perrelli	239

Presentazione

Paolo Mastandrea

(Università Ca' Foscari Venezia, Italia)

Il motivo primo di questo incontro si identifica col suo scopo: riunire insieme e far dialogare studiosi di svariati campi disciplinari e formazione disomogenea, però portati a condividere un giudizio ampiamente positivo sulla cosiddetta 'rivoluzione informatica' che ha investito ogni tipo di ricerca storico-letteraria e di indagine sui testi: un'evenienza che ha prodotto vantaggi enormi nelle nostre pratiche professionali, ma insieme ha costretto a rivedere metodologie solide e schemi di pensiero rigidi; trovando concentrati in taluni ambienti accademici - parliamo delle aree umanistiche - durevoli atteggiamenti di preconetto, di sospetto, quasi di insofferenza verso una novità ingrata. Ora che le prime classi dei 'nativi digitali' hanno concluso il loro ciclo di studio fino al dottorato, è il momento di tracciare un bilancio (e mi scuso in anticipo per alcuni non evitabili ricordi autobiografici), ma soprattutto di indicare le prospettive. Ci proponiamo di espandere risorse spesso già elaborate e realizzate nel corso degli ultimi decenni, perciò lo sviluppo degli archivi digitali e dei sistemi d'interrogazione, con relativa suite di applicativi, seguirà strade e orientamenti diversi, ma al fine unico di portare al massimo livello l'efficacia dell'indagine intertestuale computerizzata: dobbiamo capire se il modello è esportabile. Il nostro lavoro è comunque rivolto a una platea 'globale', senza limiti geografici e linguistici, destinato a prevedibili ricadute favorevoli sulla metodologia didattica - superiore oltre che universitaria.

Fin dalle mie aurorali esperienze con la videoscrittura, il word-procacci sing e il PC, a metà degli anni '80, ho creduto di intuire - non ero il solo, ovviamente - l'opportunità di poter accelerare al massimo qualsiasi ricerca verbale sui testi grazie a sistemi appositi di interrogazione che agivano su caratteri 'virtuali'; i *thesauri* elettronici raccoglievano insieme opere letterarie di epoca classica e postclassica, medievale e moderna, aprendo orizzonti impreveduti al tradizionale lavoro di analisi ed esegesi degli specialisti. Pur nell'impatto di un'innovazione tecnologica piovuta quasi d'improvviso sul tavolo di ogni studioso e studente di *humanities*, la facilità gratuita insieme alla leggerezza di esecuzione delle indagini anche più complesse - prima di allora sconsigliate anzitutto dai costi umani della fatica, dalla dismisura tra sforzi compiuti e risultati ottenuti: problemi ora risolti dalle macchine - modificavano d'un tratto le pratiche della ricerca.

La critica letteraria in Italia, che nella prima parte del novecento aveva assecondato volentieri tutte le istanze dell'estetica idealistica, salvo poi rincorrere dopo la morte di Croce qualunque moda venisse d'oltralpe, era finalmente pronta ad accogliere senza timore la nazionalistica esortazione alle concordanze di Contini, possedendo ormai strumenti interpretativi di estrema precisione e inimmaginabile potenza. A partire dagli anni '90 la consultazione dei CD-ROM del TLG, del PHI, della LIZ ecc., come più tardi l'accesso diretto ai grandi *corpora* trasferiti sulla rete web, si è via via imposta come pratica necessaria in ogni nostro campo disciplinare – pur tra le remore estenuanti e le resistenze passive cui accennavo.

Se veniamo all'oggi, uno dei benefici dell'informatica, e certo non l'ultimo, è di aver ridotto il solco fra scienziati e letterati che lungo un secolo si era invece allargato pericolosamente. La tecnologia digitale, mentre velocizza oltremodo le procedure d'indagine, incide pure in silenzio sulle abitudini dei ricercatori, portandoli a sperimentare nuove metodiche, semplicemente a mettere il naso dove era difficoltoso arrivare, semmai a provare quel piacere unico che deriva dalle scoperte casuali e si chiama 'serendipità'.

Ciò tuttavia non intacca la sostanza né altera i contenuti dell'insegnamento e della trasmissione del sapere come sono stabiliti da sempre nella prassi accademica; nessun serio conflitto epistemologico sorge a insidiare le basi istituzionali delle vetuste, gloriose materie. La ricerca nei più vari campi delle scienze umane, dalla linguistica all'archeologia, dagli studi letterari a quelli storici, filosofici, giuridici, artistici, musicali e così via, può servirsi di sofisticate risorse digitali; proprio da connubi (talora, semmai, da incontri poco più che casuali) fra *scienziati* e *umanisti*, al fine di individuare nel corso delle rispettive indagini ogni forma di automatismo utilmente trattabile con mezzi elettronici, sono venute spinte all'impiego competente degli strumenti, senza che in alcun modo fossero toccate le basi delle discipline specifiche che se ne avvalgono.

Qualunque *testo* (non solo letterario, non solo antico, non solo scritto in greco o in latino) è privo di senso fuori da una realtà storica, cioè da un *contesto* documentale; di qui un'esigenza per noi prioritaria di mettere a frutto nei singoli campi le risorse telematiche comuni, di accedere contemporaneamente a tutte le biblioteche e archivi digitali: nella forma non solo di strumenti tecnici funzionali, ma di prodotti di metodi euristici elaborati e dunque di universale avanzamento scientifico. Per quanto ci riguarda, è auspicabile si sviluppino attività di *digital studies* che sappiano compiere esplorazioni sull'intero sistema letterario nella prospettiva dello sviluppo storico, al fine di individuare le caratteristiche morfologiche dei singoli testi. La peculiarità dell'Informatica umanistica può allora indicarsi in una *variazione di scala* dei problemi affrontabili e risolvibili con l'applicazione delle sue metodiche; per fare un esempio, se la filologia tradizionale abilita a uno studio fine del dettato formulare presente in uno o più esametri di Omero o di Virgilio, la macchina permette in tempi

brevissimi il reperimento e l'esame di tutte le formule ricorrenti nei poemi epici fino a Nonno e Corippo; ancora, un'indagine intertestuale su qualsiasi campo linguistico-letterario (greco o latino, italiano o inglese o persiano) oggi si effettua con una rapidità di esecuzione e una certezza di risultato inimmaginabili, di modo che il critico è posto in grado di affrontare problemi di analisi fine sopra repertori documentali che trova preselezionati e ordinati sullo schermo.

A monte c'è sempre l'esigenza di una base oggettiva e sicura di conoscenza dei dati che può farsi risalire all'insegnamento della *Quellenforschung* positivista; e per tale via, facendo tesoro della necessaria distinzione tra riprese formali e raffronti tematici che da un'opera all'altra si trasmettono con modalità cui far corrispondere strumenti di lavoro almeno in parte diversi, sarà possibile riannodare il legame tra studio letterario e metodo scientifico. Certo è latente il rischio di perdere in profondità ciò che si guadagna in estensione, ma l'eventuale smarrimento di dettagli è comunque ben ricompensato dall'ampiezza dell'angolo visuale, oltre che dalla raggiunta concordia negli scopi (da sempre apparsa vantaggiosa, e ormai irrinunciabile) tra informazione tecnologica e sapere *umanistico* – cioè rivolto, se ci si perdona l'enfasi, all'interesse primario dell'umanità.

La scelta di campo a favore della filologia digitale, cioè dell'analisi dei dati su base 'quantitativa', si configura come un guardare avanti con fiducia collettiva, così sottraendo definitivamente la disciplina ai soggettivismi di ogni estetica idealistica o altrimenti ideologica; agli esclusivismi pregiudiziali di chi valuta la *critica del testo* (cioè l'attività individuale volta alla sua 'costituzione') esercizio poizore e sublime rispetto alla ricerca il minimo elemento per descrivere la 'storia della tradizione'. Non ho mai cessato di pensare che l'ordine della dittologia dei termini nel titolo scelto da Giorgio Pasquali¹ fosse casuale. Muovendo così i giovani a rifondare nella nostra inquieta cosmopoli la *New Republic of Letters* di cui parla in un suo bel libro Jerome McGann.²

Dati i tempi dilatati della stampa di questi Atti, che giunge al traguardo più di tre anni dopo il Convegno, i partecipanti hanno potuto sfruttare le pagine finali del volume per aggiungere eventuali dati di informazione ulteriore o di bibliografia recente. D'altro lato, è la stessa dinamicità delle edizioni e degli archivi online ove i materiali di cui si parla qui sono solitamente presentati che ci fa rimandare i lettori alla consultazione diretta dei singoli siti.

1 Pasquali, Giorgio [1934] (1952). *Storia della tradizione e critica del testo*. Firenze: Le Monnier.

2 McGann, Jerome (2014). *A New Republic of Letters. Memory and Scholarship in the Age of Digital Reproduction*. Cambridge (USA): Harvard University Press.

Considerazioni introduttive

Nicola Palazzolo

(Università degli Studi di Perugia, Italia)

1.

Sono trascorsi tre anni dal convegno di Firenze, *Diritto romano e scienze antichistiche nell'era digitale*,¹ nel quale si sono gettate le basi, se non di una collaborazione, quanto meno di una conoscenza tra ricercatori che – su sponde disciplinari diverse, conseguenza di una sciagurata divisione in rigidi settori scientifico-disciplinari – condividono però non solo lo stesso oggetto di studio, l'Antichità classica, e le stesse fonti, esaminate da punti di vista diversi, ma anche gli strumenti tecnologici utilizzati, le tecniche di reperimento e di trasmissione delle fonti, gli strumenti informatici per l'analisi dei testi, la ricostruzione delle parti mancanti, l'attribuzione di paternità.

Oggi siamo qui per vedere se è cambiato qualcosa, se si è fatto qualche passo in avanti, oltre quello, innegabile, di una maggiore consapevolezza della necessità di standard comuni, e quindi trasferibili ad altre ricerche e ad altre fonti, che può farci uscire finalmente da quella sorta di condanna che si portano dietro le scienze umanistiche, quella concezione proprietaria che impedisce qualsiasi cooperazione tra i vari sistemi.

Lo scopo del convegno è duplice: dar conto delle due ricerche PRIN, che hanno dato occasione all'organizzazione del convegno, quella che fa capo all'amico Paolo Mastandrea, e quella che moralmente fa capo ancora a me, ma che per ragioni anagrafiche è stata coordinata negli ultimi anni dal mio allievo Francesco Arcaria. Dar conto allora di queste due ricerche non tanto dal punto di vista dei risultati ottenuti nei due specifici ambiti disciplinari, ma piuttosto da quello delle metodologie digitali utilizzate per raggiungere questi risultati. Accanto a questo primo obiettivo il secondo è verificare se l'auspicio di tre anni fa, appunto di un inizio di collaborazione tra studiosi che spesso neppure si conoscevano, pur avendo notevoli punti di convergenza, ha avuto un seguito, nel senso che si è iniziato un percorso comune tra ricercatori provenienti da discipline apparentemente diverse.

1 Gli Atti del convegno sono stati pubblicati in *Diritto romano e scienze antichistiche nell'era digitale = Convegno di studio* (Firenze 12-13 settembre 2011). Torino: Giappichelli, 2012.

Per il primo profilo, e limitandomi alla ricerca da me guidata, BIA-Net (Bibliotheca Iuris Antiqui in rete) è un risultato che va oltre l'utilità specifica per chi fa ricerca nel diritto romano. Non solo si è prodotto uno strumento di cui possono servirsi tutti gli studiosi di scienze antichistiche (testi filologicamente controllati e uniformati all'edizione critica comunemente accettata, bibliografia multidisciplinare e dotata di sofisticati strumenti semantici, un Thesaurus che si arricchisce continuamente - già oggi conta oltre 10.000 termini controllati, articolati in un albero classificatorio multilingue).

Tutto ciò sarebbe stato possibile anche con i CD-Rom, come lo avevamo fatto finora, ma lo si è fatto in linea, utilizzando tutte le codifiche e gli standard propri delle tecnologie digitali (XML come linguaggio di trasmissione, TEI per la struttura dei dati, Unicode per i caratteri), quindi standard aperti e dati esportabili e riutilizzabili in altri ambiti per ricerche diverse. È stata dura, sono occorsi parecchi anni di lavoro (rispetto all'ultima edizione su CD-ROM sono passati 12 anni, scontando le richieste pressanti e i malumori di chi in tutti i convegni ci chiedeva quando sarebbe uscita la nuova edizione), ma oggi possiamo dire con soddisfazione di aver resistito alle sirene, e di avercela fatta. Rimangono da risolvere alcuni problemi organizzativi relativi alla diffusione dei dati, ma penso che nei prossimi mesi potremo chiudere il lavoro. Daria Spampinato e Marco Giunti entreranno nei particolari relativamente a due degli aspetti della ricerca, che vanno proprio nella direzione dell'interoperabilità con altri archivi digitali.

Per il secondo profilo già l'organizzazione del convegno credo sia un notevole risultato. La collaborazione tra me, con i miei allievi sparsi per l'Italia, e Paolo Mastandrea con la sua scuola, con cui ci siamo conosciuti proprio a Firenze in quel primo convegno, è il frutto più tangibile di quello spirito che ci animò allora. E basta una scorsa al programma del convegno per accorgersi del passo avanti che è stato fatto: non più una vetrina delle cose fatte, allo scopo di farsi conoscere, ma una messa a fuoco del problema della collaborazione nelle sue varie sfaccettature.

Sul piano più strettamente dei progetti di ricerca un primo tentativo è stato fatto con il progetto europeo coordinato da Maurizio Lana di un corpus globale virtuale di testi latini digitalizzati, dalle origini al Rinascimento, in pratica una *digital library* del latino, nella quale si interrogano i diversi database come se fossero uno solo e si accede a tutti da un'interfaccia unica. Avevano dato la loro disponibilità al progetto diverse basi di dati: *DigilibLT* anzitutto, con i testi letterari in prosa latina dal II al VII sec. d.C.; *Musisque Deoque*, con la poesia latina dal III sec. a.C. al Rinascimento; la *Bibliotheca Iuris Antiqui*, con il corpus completo di testi giuridici latini e greci dalle origini a Giustiniano; e infine DIGIMED, con i testi medievali di argomento storico-geografico, insomma una collezione già così pressoché completa (circa 13 milioni di parole), e con prospettive di ulteriore allargamento, del patrimonio testuale della latinità.

Il progetto è stato presentato per due volte in occasione di bandi europei, e si è collocato in buona posizione, anche se alla fine non è stato finanziato. Io vorrei spingere a riprovare ancora, magari correggendo e migliorando il progetto nei punti che appaiono più deboli, ma a non rinunciare, perché la strada è questa. E comunque, al di là del finanziamento europeo, la strada della collaborazione e di integrazione tra i vari progetti di digitalizzazione si impone, e richiede da noi qui ed oggi, iniziative concrete che vadano in questa direzione. Proprio per questo abbiamo affidato la relazione di apertura al prof. Gregory Crane, che a livello mondiale dirige la più importante iniziativa di collaborazione per le lingue classiche, che è il Perseus Project, e che voglio ringraziare per la disponibilità, ma abbiamo pure studiosi come Franco Niccolucci o Silvia Orlandi, che stanno lavorando all'interno di gruppi di lavoro che fanno capo a Europeana, la grande biblioteca digitale europea.

2.

Quando parliamo di strumenti per la ricerca non possiamo ignorare che i principali strumenti di cui ci serviamo per la ricostruzione delle civiltà classiche sono i documenti scritti, sono quelli che chiamiamo 'testi'.

Abbiamo presente certamente che tutta una serie di altre ricerche si basa su documenti non testuali, della cosiddetta 'cultura materiale' (prendiamo ad esempio tipico l'archeologia), e che di questo tipo di documenti dobbiamo pur occuparci quando parliamo di applicazioni informatiche. Ma non c'è dubbio che la massa dei dati che noi trattiamo sono dati testuali, contenuti in supporti diversi (manoscritti, iscrizioni, papiri, ostraka, monete, ecc.), ma pur sempre testi.

E però occorre ripetere qui qualche distinzione, come ho già fatto nel precedente convegno, ma che voglio ribadire perché mi sembra non ancora chiara sino in fondo.

Cos'è un testo? È una successione di segni grafici, una successione di parole, è un oggetto fisico (libro, manoscritto, papiro) riprodotto in una serie di immagini digitali, o è un contenuto che l'autore ci vuole trasmettere, e che vogliamo analizzare a fini di studio e di ricerca? Sono tante le accezioni che possiamo dare all'espressione, e certo non voglio invadere il mestiere di altri facendo il tentativo di dare una risposta a questa domanda. Ma è chiaro che ciascuna di quelle accezioni cui prima accennavo dipende strettamente dal tipo di utilizzo che se ne intende fare.

Dico questo perché nell'ultimo decennio il tema della digitalizzazione del patrimonio testuale dell'umanità, attraverso la creazione di vere e proprie 'biblioteche digitali' si è progressivamente imposto come uno dei più significativi prodotti dell'applicazione delle nuove tecnologie digitali. C'è stato un grosso *battage* pubblicitario, anche commerciale, sulla velo-

cità di digitalizzazione dei grandi patrimoni bibliotecari. Ma spesso per i non addetti ai lavori una cosa non è chiara, che spesso invece è il limite principale di queste iniziative: il fatto che l'opera digitalizzata sia quasi sempre il supporto fisico, il libro inteso come bene culturale, da proteggere e conservare come tale, nella sua fisicità, non il suo contenuto, da mettere a disposizione dei ricercatori, e capace quindi di essere utilizzato all'interno di un'attività di ricerca. E ciò perché, nella gran parte dei casi, quando si parla di digitalizzazione di grandi patrimoni testuali, quasi sempre si tratta di immagini fotografate attraverso sistemi più o meno avanzati, magari con un sistema di indici che rimandano alle singole pagine in cui quell'argomento è trattato, ma senza che il testo in quanto tale possa essere trattato e analizzato sia dal punto di vista linguistico che da quello filologico.

Per rimanere dentro il mio settore disciplinare è stato con viva curiosità che sono andato a guardare nel sito dell'Università di Bologna un oggetto che, sotto il pomposo titolo di *Corpus Iuris online*² proponeva nient'altro che una collezione di immagini tratte da una prestigiosa edizione della Glossa accursiana, corredata da indici di riferimento: una bella iniziativa, non c'è che dire, sarà anch'essa una biblioteca digitale ma tutti comprendiamo che è cosa diversissima da un testo digitale, che può essere non solo ricercato, ma analizzato, manipolato a fini diversi.

Occorre che ci diciamo che ormai si tratta di strumenti certamente utili, ma utili a scopi diversi da quelli dell'analisi testuale. La biblioteca digitale di opere in formato immagine non può essere l'obiettivo principale da raggiungere, ma solo un primo passo, sul quale poi fondare un'edizione digitalizzata non di un'immagine, ma di un testo scritto, che sia agevolmente catturabile e manipolabile dal ricercatore stesso. In questa direzione sta appunto lavorando l'*Open Greek and Latin Project* dell'Università di Lipsia.

3.

Dicevo che occorre andare avanti verso l'interoperabilità dei nostri archivi e mettere in comune i nostri sforzi per raggiungere questo obiettivo. Abbiamo un patrimonio digitale notevolissimo, ma non tutto è allo stesso livello di elaborazione. Allora lo sforzo deve essere duplice:

- portare tutti i testi che abbiamo a uno standard comune, sia dal punto di vista formale (XML, TEI, ecc.), sia da quello sostanziale (correttezza filologica attestata dalla conformità a un'edizione critica accettata). So che è difficilissimo, che richiede tanta fatica e tempi non brevi.

2 URL <http://amshistorica.unibo.it> (2017-10-19).

Richiede il lavoro dell'informatico per la conversione del formato, ma richiede anche il lavoro dello specialista, che scelga l'edizione di riferimento e controlli il testo parola per parola dando il giusto valore ai segni diacritici utilizzati in quell'edizione per uniformarli a quelli dello standard prescelto. È il lavoro che abbiamo fatto per BIA-Net sui testi giuridici latini e greci: non solo il *Corpus Iuris giustiniano*, del quale ci sono in linea tantissime versioni anche testuali, ma delle quali si ignora quale sia l'edizione di riferimento, e quindi non attendibili, ma anche per tutte le fonti giuridiche pregiustiniane, nonché per le fonti greche, anch'esse uniformate agli standard TEI (e devo ringraziare per queste l'apporto che ci è stato dato dall'Istituto di Linguistica computazionale del CNR di Pisa, sia per il testo greco delle *Novellae*, sia per quello delle costituzioni greche del *Codex Iustinianus*, entrambi testi del tutto assenti online). Lo stesso discorso che facciamo per le fonti dobbiamo farlo per le basi di dati bibliografiche. È oggi ormai un risultato acquisito la possibilità di interrogazione contemporanea di molte basi di dati, e proprio nelle ricerche interdisciplinari, quali ormai sono e dovranno essere sempre più le nostre, il vantaggio è quanto mai evidente, ma anche qui i risultati possibili sono tanto più soddisfacenti in quanto tali basi di dati hanno conosciuto in fase di analisi concettuale e progettuale una normalizzazione del linguaggio di rappresentazione. Ciò significa per i produttori di banche dati bibliografiche la necessità di convertire in un linguaggio standardizzato (XML) tutti i metadati descrittivi delle singole unità bibliografiche presenti nella banca dati.

- Rendere esportabili e interoperabili le nostre banche dati, cioè garantire il riuso dei materiali digitalizzati senza vincoli particolari. Abbandonare quindi ogni residua concezione proprietaria, e affidarci interamente a tecnologie *open source* e *open access*, ovviamente con un giusto riconoscimento, e laddove possibile un'equa remunerazione, del lavoro svolto.

La prima sessione del convegno dovrà servire a farci capire a che punto siamo in questa strada.

4.

Ma mettere in comune le risorse non è solo un problema tecnico e di standard. Il passaggio successivo è quello di avere strumenti più raffinati che ci consentano di reperire materiali che con i normali strumenti interni alle banche dati non si riuscirebbe a ottenere. Ecco che si apre tutto un campo di ricerca nel quale – almeno per restare al nostro campo – siamo molto indietro. Vi sono – è vero – diversi strumenti di cui ci serviamo ai fini della

ricerca nelle nostre banche dati, ma sono per lo più strumenti creati specificamente per un corpus spesso molto settoriale. Occorre chiedersi invece se non sia possibile andare alla ricerca di uno strumento che abbia da un lato sufficiente flessibilità per essere adattato alle esigenze di ricerca in uno specifico ambito scientifico, ma dall'altro che sia in grado di fare una ricerca ad ampio spettro entro corpora eterogenei, quali sono i nostri. Il web semantico e le ontologie dovrebbero andare in questa direzione. Cito solo l'esperienza di BIA-Net, nella quale – ed è stato uno dei punti di forza del progetto – si è creata un'apposita unità di ricerca (fatta non di giuristi né di informatici o di documentaristi, ma di filosofi, per la precisione di Logici) per studiare le possibilità che il vecchio *Thesaurus* di BIA (che pur si giovava di un patrimonio di oltre 10.000 termini strutturati entro una classificazione a 6 livelli di approfondimento) potesse non solo arricchirsi di nuovi termini ma incrementarsi con nuovi legami per essere utilizzabile nell'ambito di qualunque delle altre discipline antichistiche. Ma si parlerà anche di altri strumenti: Wordnets per le lingue classiche, ad esempio, è uno degli strumenti semantici che potrebbero essere utilizzati in tutti gli ambiti disciplinari delle scienze del mondo antico.

Ecco perché si è voluto dedicare un'apposita sessione del convegno agli strumenti semantici.

5.

Peraltro la ricerca informatica in antichistica non è certo limitata alla creazione di corpora digitali più o meno ampi e agli strumenti avanzati per ricercare le occorrenze di un termine o di una frase nell'immenso corpus delle fonti antiche, o per ricercare la bibliografia collegata a un testo o anche solo a una frase. Dobbiamo andare oltre: e c'è qualcuno che già lo fa. Ad un certo punto sorge l'esigenza di poter, attraverso il computer, compiere non più una ricerca *dei* testi (per esempio, quelli che contengono un nome, una parola, o una sequenza di parole: le cosiddette concordanze), ma una ricerca *sui* testi, appunto l'analisi dei dati.

Allora vediamo fino a che punto le metodologie e anche gli strumenti tecnologici che alcuni di noi stanno sperimentando possono essere utilizzati anche da altri. Prendiamo il caso dell'attribuzione di paternità di un testo, prendiamo quello della ricostruzione di testi frammentari (generalmente papiri e iscrizioni), o infine quello della creazione di nuove edizioni critiche.

Quello dell'attribuzione di un testo disponibile nel mondo digitale è per esempio un campo nel quale molti studiosi potrebbero ritrovare più di un motivo di interesse. Nel precedente convegno a Firenze ha colpito la relazione di Maurizio Lana, che ha mostrato come, sulla base di metodi quantitativi, si possa tentare di individuare l'autore materiale di testi

legislativi imperiali in funzionari che abbiano ricoperto cariche di rilievo nella cancelleria imperiale, e di cui siano noti altri scritti.

La terza sessione del convegno sarà appunto dedicata a queste problematiche.

6.

E infine una parte che apparentemente è meno legata agli studi antichistici, ma che in realtà ci riguarda tutti. Quali le modalità organizzative e le forme giuridiche per un'impresa del genere, che non sia solo una mera attività di ricerca all'interno dei nostri studi e laboratori, ma si voglia proporre anche sul mercato? Quali i beni da tutelare e quali quelli da lasciare in uso libero da vincoli alla comunità scientifica? E da che cosa tutelarli? Quali i pericoli da cui difendersi? Gli attacchi degli hacker o la copiatura e l'esportazione abusiva dei dati? Tutela dal plagio o dalla contraffazione?

Da un altro punto di vista: si è parlato di tendenza all'open access; ma come fare a remunerare l'eventuale plus-valore che i metadati creati dall'attività dei ricercatori aggiungono al documento, che considerato nella sua essenzialità deve essere invece di pubblico dominio e ad accesso gratuito? Le due relazioni della quarta sessione cercheranno di rispondere a questi interrogativi.

7.

Chiuderà il convegno una Tavola Rotonda, nella quale vorremmo fare il punto della situazione su tutti questi problemi che, insieme, costituiscono la Biblioteca digitale, non un semplice corpus digitalizzato di documenti, un mero deposito nel quale andare a ricercare ciò che serve, ma piuttosto un complesso organizzato di servizi finalizzati all'utente-ricercatore. È possibile creare uno strumento simile per le scienze dell'Antichità, uno strumento che metta insieme corpora distinti e di differenti tipologie, anche se accomunati dagli stessi standard, ma anche strumenti adatti per le diverse tipologie di ricercatori, e organizzati in una struttura istituzionale, pubblica o privata, che ne consenta lo sfruttamento anche all'esterno delle comunità di ricerca che l'hanno creata? Io spero di sì, ed è questo il compito che vorrei affidare alle generazioni che ci seguono.

Il nostro, come dicevo, non vuole essere un convegno di presentazioni di cose fatte, una passerella di progetti realizzati o da realizzare. Ci saranno anche queste, ed è giusto farle conoscere, ma deve essere solo il punto di partenza. Il convegno deve essere anzitutto un confronto di idee, di proposte, di cose da fare insieme. Io mi auguro che da questo convegno

escano fuori i nuovi progetti per le scienze dell'Antichità, su cui cimentarci insieme nei prossimi anni.

Noi abbiamo fatto il primo passo, che forse era il più difficile: abbiamo fatto almeno il tentativo di dialogare tra noi, di scambiarci le esperienze. Adesso è il tempo di agire, e di agire insieme, perché le risorse pubbliche sono sempre di meno e quelle private per un campo di studio come il nostro richiedono menti illuminate che al momento non vediamo all'orizzonte.

Parte 1.

Costruzione e gestione
di corpora e piattaforme digitali

Standard e interoperabilità degli archivi digitali archeologici: il progetto ARIADNE

Franco Niccolucci

(Università degli Studi di Firenze, Italia)

Abstract The increasing use of digital documentation has created an uncontrolled growth of archaeological datasets, determining a fragmentation that makes data reuse difficult. Starting from this consideration, the European project ARIADNE has undertaken a research of methods and tools to overcome linguistic, administrative and data organisation barriers that prevent the integration of datasets related on location, period or subject. Such research aims at creating a registry of online accessible archaeological resources, enabling resource discovery via a unified portal. Such research functionality will enable facet queries, i.e. by theme, period or location. Moreover, the project is producing a number of services that improve data access, for example the visualisation of graphical information (images, 3D models), and develop specific documentation standards based on international standards such as the CIDOC CRM.

Sommario 1 Introduzione. – 2 ARIADNE: ambito di azione e obiettivi. – 3 La Resource Discovery in ARIADNE. – 4 La standardizzazione. – 5 L'interoperabilità dei database. – 6 L'applicazione delle tecniche di data mining. – 7 Prospettive e problemi. – 8 Riconoscimento.

Keywords Archaeology. Research infrastructure. Resource discovery. Integration. Standards.

1 Introduzione

Il progetto ARIADNE (*Advanced Research Infrastructure for Archaeological Data Networking in Europe*, cf. ARIADNE 2015) fa parte del gruppo numeroso di progetti finanziati dall'Unione Europea per costituire infrastrutture di ricerca, cioè sistemi integrati di risorse umane e tecnologiche che operano in tutti i campi scientifici, offrendo ai ricercatori servizi avanzati e collegando, integrandoli fra loro, analoghi sistemi esistenti a livello nazionale o regionale. Si tratta di progetti sostanzialmente diversi da quelli tradizionali, in cui i partecipanti collaborano per il raggiungimento di un obiettivo specifico di ricerca. Nel caso delle infrastrutture di ricerca, invece, l'obiettivo del progetto è generale e di base, correlato all'ambito disciplinare ma svincolato da un tema circoscritto e determinato. Se quindi la caratteristica di un progetto di ricerca è la sua temporaneità - una

volta raggiunto il risultato, o terminato il tempo a disposizione, il progetto termina - un'infrastruttura di ricerca ha l'ambizione di essere durevole, continuando a funzionare anche oltre il periodo del finanziamento, che serve soltanto per metterla in piedi a livello europeo.

Simili sistemi sono caratteristici delle scienze fisiche e naturali, dove sono spesso necessarie strumentazioni complesse e di grandi dimensioni oppure è auspicabile realizzare sistemi complementari compatibili, e in cui la condivisione dei risultati e il lavoro di équipe sono strumenti indispensabili per l'avanzamento della conoscenza. Sembrerebbero invece meno congeniali al modo tradizionale di lavorare nelle scienze umane, basato sulla ricerca individuale del singolo ricercatore. Tuttavia, la presenza sempre più diffusa di metodologie basate sull'informatica e la disponibilità di corpora digitali rendono anche le discipline umanistiche candidate idonee alla creazione di infrastrutture di ambito europeo. Infatti, il meccanismo previsto dall'Unione Europea prevede l'individuazione delle tematiche e delle relative comunità di ricercatori ormai mature per sperimentare la creazione di un'infrastruttura europea, in genere partendo da sistemi già esistenti a livello nazionale. Un ciclo di quattro anni, spesso ripetuto con un'ulteriore prosecuzione di altri quattro, porta queste comunità all'integrazione e, successivamente, all'autosufficienza. Da questo punto in poi l'infrastruttura dovrà essere un soggetto autonomo con proprie risorse. Un certo numero di tematiche vengono periodicamente individuate come prioritarie dall'ESFRI (*European Strategy Forum on Research Infrastructures*, un comitato di esperti nominati dagli stati membri), sia per la loro importanza che per la disponibilità di gruppi di stati a finanziarle in modo permanente. Le infrastrutture corrispondenti si costituiscono in enti di diritto europeo chiamati ERIC (*European Research Infrastructure Consortium*) di solito partecipati e con un contributo annuo da parte dei ministeri nazionali della ricerca degli stati che vi partecipano. Nel campo delle scienze umane esiste, ad oggi, un solo ERIC, denominato DARIAH (*Digital Research Infrastructure for the Arts and Humanities*), che ha come oggetto sociale «migliorare e supportare la ricerca e l'insegnamento basati sul digitale in tutte le scienze umane e le arti». Esistono invece varie infrastrutture di ricerca nel campo delle scienze umane, una delle quali è appunto ARIADNE.

2 ARIADNE: ambito di azione e obiettivi

ARIADNE nasce dal riconoscimento da parte della Commissione Europea dell'importanza crescente per la ricerca archeologica degli archivi digitali. Inserita nel 2012 in una *'long list'* di circa trenta nuove tematiche potenziali soggetti di finanziamento, dall'ambiente marino all'astrofisica, poste in concorrenza fra di loro a prescindere dall'ambito disciplinare,

l'archeologia digitale è riuscita ad affermarsi in questa forte competizione che ha visto solo pochissime delle proposte conseguire il successo. Il progetto è quindi iniziato nel Febbraio 2013 con un finanziamento quadriennale di circa 8 milioni di euro. Vi partecipano 24 partner da 15 paesi europei, in genere centri di importanza nazionale come le Accademie delle Scienze, i Consigli Nazionali delle Ricerche e centri accademici di livello internazionale. Per l'Italia sono presenti il PIN - Polo Universitario di Prato (coordinatore), il CNR e il Ministero dei Beni e Attività Culturali e del Turismo, oltre all'AIAC, Associazione Internazionale di Archeologia Classica con sede a Roma ma operante in tutta l'area mediterranea. Come risulta dal nome del progetto, l'obiettivo principale è l'integrazione degli archivi archeologici digitali.

La necessità di un progetto d'integrazione nasce dalla constatazione (Niccolucci 2013) del ruolo e dell'importanza crescente assunta dalla documentazione digitale negli ultimi venti anni. Se negli anni '90 la creazione di un database di scavo era ancora un caso raro e spesso il soggetto di una pubblicazione scientifica, questo tipo di attività è ormai diventata una pratica comune e non fa più notizia: fa anzi scalpore la sua mancanza in un'indagine archeologica. Ciò ha prodotto una grande quantità di dati organizzati in modo estremamente eterogeneo, e a sistemi spesso incompatibili tra di loro. L'uso esteso della fotografia digitale e, in tempi più recenti, delle immagini tridimensionali ha prodotto (Niccolucci, Richards 2013) un aumento esponenziale della dimensione media degli archivi e quindi difficoltà di gestione in mancanza di idonei strumenti software. Tralasciando gli archivi di proprietà e accesso esclusivo di ricercatori ed enti di ricerca, è sempre più presente la pubblicazione online dei dati, nello spirito di condividere i risultati delle proprie ricerche, finanziate con risorse pubbliche e quindi non di proprietà esclusiva di un ricercatore o di un ente. Sull'accesso pubblico a tali archivi con sistemi online si è da tempo aperto un dibattito sui limiti della proprietà intellettuale di tali dati, favorito anche da taluni eccessi nella protezione che hanno visto ritardarne la pubblicazione a lungo, e in qualche caso estremo anche per sempre. Perché costretti dalle clausole del finanziamento del progetto di ricerca, per convinzione personale o per seguire un filone sempre crescente, moltissimi ricercatori hanno comunque reso disponibili i propri dati su Internet, facilitati anche dalla semplificazione tecnica ormai raggiunta dai meccanismi di pubblicazione online di database. Ciò corrisponde allo scopo della Dichiarazione di Berlino sull'Accesso Libero alla Conoscenza nelle Scienze e nelle Discipline Umanistiche (cf. Berlin-Declaration 2015), sottoscritta sin dal 2005 da tutti gli atenei italiani e successivamente anche dal CNR e dagli altri enti pubblici di ricerca, che impegna gli enti sottoscrittori (e di conseguenza anche il loro personale) a rendere disponibili i risultati delle ricerche prodotti in qualunque forma, siano essi articoli scientifici, dati grezzi e relativi metadati, fonti, rappresentazioni digitali

di materiali grafici e multimediali, e così via. Tali risultati dovranno essere resi disponibili online con diritto di copia, distribuzione e rielaborazione, con il solo obbligo di citazione della fonte. In sintesi, ogni ricercatore italiano (e, in generale, europeo) deve rendere disponibili online e senza limitazioni d'uso i risultati della propria ricerca e i dati grezzi utilizzati o prodotti nel corso di questa. Per quanto riguarda gli archivi amministrativi, considerazioni di privacy e di sicurezza ne potrebbero limitare l'accesso, ma ciò riguarderebbe solo una frazione ristrettissima d'informazioni, peraltro spesso prive d'importanza scientifica. Queste limitazioni possono essere facilmente implementate in sistemi informatici senza pregiudicare l'accessibilità alla stragrande maggioranza delle altre informazioni. Questo elimina anche la scusa principale per mantenere non accessibili gli archivi, fino a non molto tempo fa diffusa tra funzionari tanto conservatori quanto ignoranti delle funzionalità della tecnologia informatica.

Dunque il panorama che si offre al ricercatore vede un gran numero di archivi accessibili online, però eterogenei e con strutture e organizzazione incompatibili, e spesso di dimensioni significative: si tratta del lavoro accumulato da ricercatori, équipes e istituzioni che hanno generato dati sia come risultati di ricerca che a seguito di attività di gestione amministrativa. Ciò costituisce il terreno ideale per proporre l'integrazione

Occorre precisare il significato da attribuire a quest'ultimo termine. Ci si potrebbe infatti chiedere che senso abbia allineare archivi che riguardano, per esempio, sepolture dell'Età del Ferro in Norvegia e insediamenti fenici in Spagna. Ovviamente la domanda ha un senso, e risposta negativa, in esempi estremi come il precedente. Integrazione significa, invece, rendere compatibili archivi digitali che già lo sono riguardo al contenuto, per i quali cioè l'integrazione ha un senso prima di tutto archeologico, per pertinenza di argomento, di periodo o di copertura geografica. Integrazione non significa in ogni caso creare un super-archivio in cui far confluire tutti quelli esistenti, che devono essere mantenuti dove si trovano ma devono essere collegati da un sistema comune di schedatura e accesso. Si deve dunque raccoglierne le informazioni descrittive essenziali, i metadati, con cui creare un elenco intelligente, un 'pubblico registro'. Questo dovrà contenere informazioni sul loro contenuto e la loro organizzazione, e, soprattutto, permettere di accedere direttamente al loro contenuto là dove questo è conservato. Dunque la creazione di questo catalogo intelligente è il primo obiettivo da perseguire. A questo riguardo, ARIADNE ha sviluppato un proprio *Registry* dove sono censiti gli archivi archeologici europei, partendo, in una prima fase, da quelli dei partner del progetto, che costituiscono comunque una percentuale molto elevata, per non dire la quasi totalità di tutti quelli esistenti. È previsto in un secondo momento di aprire il Registro a tutti i produttori di dati in modo da arrivare nel medio periodo a un'anagrafe totale. L'apertura al pubblico del Registry all'inizio del 2016 ha realizzato il primo obiettivo del progetto: integrazione degli

archivi in un unico registro europeo. Oltre ai partner dei quindici paesi partecipanti si sono aggregati al progetto istituzioni nazionali di altre cinque nazioni ottenendo una copertura quasi completa dell'Europa. Il Registro ha così realizzato il primo dei servizi programmati: la *Resource Discovery* o Individuazione delle Risorse.

3 La Resource Discovery in ARIADNE

Il servizio di *Resource Discovery* permette di ricercare un archivio archeologico fra quelli censiti nel *Registry*. Attualmente il *Registry* comprende i metadati di circa 2.000.000 archivi, assai eterogenei fra di loro. Si tratta infatti di singoli file, cartelle con file di testo, immagini e altri dati, database strutturati e così via. Come già detto, si tratta soltanto di una parte (quasi completa) dei metadati degli archivi gestiti dai partner. A fronte di queste dimensioni, il semplice censimento non risolve però i problemi del ricercatore interessato a una specifica questione. ARIADNE fornisce gli strumenti per individuare nel *Registry* le risorse pertinenti all'interesse del ricercatore che vi accede. Questa ricerca può essere fatta per parole chiave sulla descrizione dell'archivio oppure per *facets*, cioè basata su una classificazione degli archivi secondo soggetti predeterminati che nel caso specifico sono il periodo di riferimento, l'area geografica e la tipologia dei contenuti. In questo modo è possibile ricercare, ad esempio, gli archivi che contengono dati relativi al VI secolo a.C., oppure relativi all'area mediterranea o, infine, i soli archivi relativi a necropoli; oppure incrociare fra di loro tali criteri di selezione. È previsto anche l'uso di una *time line*, cioè di un asse temporale che individua visivamente l'intervallo temporale desiderato. È, infine, possibile visualizzare su una mappa gli archivi relativi a un'area specifica; in questa modalità, tuttavia, si potrebbero trascurare archivi che coprono un'area più vasta, come quelli nazionali, perché questa selezione su base geografica è significativa quando l'archivio riguarda un'area ristretta, mentre per archivi che coprono aree vaste, come ad esempio quello del MIBACT che copre tutto il territorio nazionale, tale informazione è invece rilevante solo a livello di record o di gruppi di record.

4 La standardizzazione

Un aspetto importante di ARIADNE riguarda la standardizzazione dei sistemi di documentazione archeologica. È infatti questo un passaggio preliminare all'integrazione e ancor di più all'interoperabilità, cioè alla 'fusione' virtuale del contenuto di archivi digitali distinti, che richiedono di essere ricondotti a una struttura comune. Per questo scopo è stato individuato il sistema CIDOC CRM (CIDOC-CRM 2015), un modello concettuale che

definisce una struttura formale per descrivere i concetti impliciti ed espliciti e le relazioni usati nella documentazione del patrimonio culturale. Si tratta dello standard ISO 21127:2014 ormai generalmente accettato come 'lo' standard e adottato da prestigiose istituzioni come il British Museum, l'Istituto Archeologico Germanico e molte altre.

A seguito della crescita spontanea dei sistemi di archiviazione digitale dei dati archeologici, sono stati utilizzati nel tempo sistemi diversi e incompatibili fra di loro per strutturare le informazioni. Anche quando è stato sviluppato nei primi anni 2000 il CIDOC CRM su iniziativa del CIDOC (Comitato Internazionale per la Documentazione di ICOM), l'associazione internazionale dei musei, la sua adozione è stata assai faticosa. Si è a lungo sostenuto (e ancora oggi talvolta accade) che fosse inutilmente complicato e di difficile applicazione. Ma, come per i compiti dei cattivi studenti, la difficoltà stava solo nella pigrizia mentale e nell'ignoranza di chi doveva utilizzarlo e non nelle qualità del sistema, che si è progressivamente affermato per il suo rigoroso apparato logico e formale.

Il CRM è alla base della costruzione di collegamenti fra sistemi di organizzazione dei dati diversi fra loro, che vengono 'mappati' sul CRM per stabilirne le corrispondenze. Un gruppo di ricercatori all'interno del progetto ha sviluppato uno standard per la gestione dei dati archeologici che è stato chiamato CRM_{Archeo}. Questa estensione del CRM aggiunge categorie e relazioni specifiche del campo archeologico, ad esempio il concetto di 'unità stratigrafica' di uso corrente nella documentazione di scavo e non presente nella versione originale del CRM. Ulteriori specializzazioni sono in corso di realizzazione per casi particolari quali ad esempio i risultati di analisi scientifiche (il noto sistema di datazione basato sulla concentrazione dell'isotopo carbonio 14 nei reperti organici ne è un esempio) usati nel campo della ricerca archeologica, che producono sia informazioni precise sui reperti che collezioni di riferimento di fondamentale importanza per la loro classificazione. Oltre al già ricordato ¹⁴C, vi sono molte tecnologie che permettono di caratterizzare i materiali e la loro provenienza, e tecniche come l'analisi del DNA che forniscono preziose informazioni sui resti umani e animali. Tutte queste tecniche si basano su repertori di riferimento, che vengono arricchiti dai risultati di ciascuna nuova analisi. Si tratta quindi di un settore in cui la disponibilità e la condivisione dei dati è, se possibile, ancora più importante

5 L'interoperabilità dei database

ARIADNE intende integrare anche i sistemi più complessi presenti in alcuni dei paesi che vi partecipano o prodotti da ricerche archeologiche. Si tratta di database che è necessario prima di tutto standardizzare, cioè allineare con un sistema standard di organizzazione dei dati che ARIADNE

ha individuato nella specializzazione del CIDOC CRM per i dati archeologici sopra descritta. Tale operazione è attualmente testata su alcuni casi relativamente più semplici per procedere successivamente a quelli di maggiore complessità. Sono infatti già stati integrati, a livello appunto sperimentale, dei database numismatici forniti da varie nazioni, creando un unico punto di accesso per le ricerche su un database distribuito. In questo caso un'ulteriore difficoltà si presenta a livello linguistico perché gli archivi sono realizzati, naturalmente, nelle varie lingue nazionali. Se questo non costituisce un problema per la *Resource Discovery*, in quanto la descrizione dell'intero archivio è compilata appositamente per il *Registry* in lingua inglese, le chiavi di ricerca in un archivio integrato multilingue devono essere formulate in una qualunque delle lingue presenti (e preferibilmente anche in altre) e automaticamente risolte in tutte le lingue, utilizzando dei thesauri multilingue. La disponibilità di un thesaurus multilingue per la numismatica, creato in un progetto precedente, ha perciò favorito la scelta di questo settore. Per gli altri si procederà gradualmente all'armonizzazione dei thesauri disponibili, restringendo la scelta dei parametri di ricerca ai termini presenti in questi.

Il rilascio dello standard definitivo proposto per la documentazione archeologica e la creazione dei primi sistemi interoperabili, avvenuti nel corso dell'ultimo anno del progetto, costituiscono un'attività da continuare in un'eventuale prosecuzione del progetto stesso.

6 L'applicazione delle tecniche di *data mining*

Molti degli archivi censiti in ARIADNE sono documenti di testo: rapporti di scavo, relazioni di interventi, documentazione testuale in generale. Per questo tipo di dati l'unica informazione strutturata disponibile è la breve descrizione fornita insieme all'archivio stesso contenuta nel *Registry*. Si tratta comunque di un'informazione limitata, quindi a livello sperimentale si è attivata una ricerca basata sul *data mining*, cioè sull'indicizzazione dei testi mediante l'individuazione di termini specifici contenuti in thesauri o liste di termini. Questo tipo di applicazione è in stato di sviluppo avanzato per la lingua inglese, in cui sono scritti circa il 30% dei testi presenti negli archivi censiti, e per l'olandese, che fornisce una quota di testi di poco minore. Un'applicazione analoga viene sviluppata per l'italiano all'interno di un altro progetto europeo. Il problema è meno rilevante per il tedesco, in quanto in questi paesi il sistema di archiviazione è organizzato in database e non come file 'sciolti' di testo, e quindi rientrano nel caso esaminato in precedenza. Si tratta comunque anche in questo caso di un'attività che si estende oltre la durata del progetto che ne vedrà solo una realizzazione a livello sperimentale, da completare anch'essa in un'eventuale prosecuzione.

7 Prospettive e problemi

Allo stato attuale, cioè fine 2017, ARIADNE ha realizzato completamente il *Registry* e il servizio di *Resource Discovery*, che quindi sono posti a disposizione della comunità di ricerca. Anche lo standard per la documentazione archeologica è in corso di pubblicazione con il necessario corredo di supporto (manuali, esempi, e così via). L'integrazione dei database, il *data mining* e altri servizi complementari sono invece disponibili solo a livello sperimentale, con una possibilità di completamento se il progetto verrà proseguito. A questo punto sarà disponibile un sistema che coprirà l'intera Europa permettendo l'accesso online a una vastissima mole di informazioni archeologiche e realizzando così un'infrastruttura digitale di dati corredata di idonei servizi di ricerca di informazioni, primo e fondamentale nucleo di un ecosistema su cui costruire un ambiente virtuale di ricerca come già esistono nel campo delle scienze biologiche, fisiche e chimiche. Si tratterà di uno strumento unico nel suo genere, di grande valore sia per la ricerca che, potenzialmente, per la conservazione, la gestione e la valorizzazione del patrimonio archeologico (Niccolucci 2015; Meghini et al. 2017).

Gli ostacoli che si frappongono al conseguimento di quest'obiettivo non sono insuperabili. A parte il reperimento delle risorse finanziarie necessarie che non sono così elevate come per altre discipline, aspetto su cui sembra inoltre esserci una sensibilità precisa nella Commissione Europea, vi sono una serie di passi che la comunità scientifica deve compiere.

La mentalità di condividere i propri dati e di favorirne il riuso è ormai un fatto consolidato e, come si è detto, un obbligo per tutti i ricercatori che appartengono alle istituzioni firmatarie della Dichiarazione di Berlino. È ormai sempre più ridotto il numero di quanti sostengono un improponibile diritto di autore a tutela dei dati prodotti dalla propria ricerca scientifica, ed è sempre più vasta l'accettazione dell'obbligo di garantirne l'accesso libero.

Si tratta quindi da un lato di rendere fluidi l'arricchimento di questa infrastruttura digitale e il suo accesso. Ciò comporta in primo luogo la creazione, dove già non esistono come nel Regno Unito, di reti nazionali di supporto per la creazione di sistemi di *hosting* degli archivi creati da singoli ricercatori o gruppi di ricerca, cui poter conferire gli archivi creati nel corso delle attività scientifiche. Per questo si devono definire procedure e politiche che da un lato ne facilitino l'integrazione, promuovendo l'uso di standard condivisi; dall'altro forniscano una 'remunerazione' in termini accademici per la creazione di tali archivi, da valutare come pubblicazioni scientifiche a tutti gli effetti. Questi sistemi di deposito dei dati potranno ovviare alle difficoltà che ancora s'incontrano a livello periferico/decentrato per pubblicare i propri database di ricerca, garantendone la qualità e la conservazione nel tempo. È sottinteso che si presuppone un livello di cultura interdisciplinare forse ancora solo parzialmente presente nel nostro paese, dove la formazione di esperti con competenze umanistiche

e informatiche è lasciata alla buona volontà dei singoli e considerata irrilevante a livello accademico, quando non dannosa per la carriera. Una volta superata la frammentazione degli archivi archeologici attraverso la loro integrazione nell'infrastruttura europea, sarà poi necessario superare la frammentazione dei meccanismi di accesso. Ancora oggi in ARIADNE, una volta individuato l'archivio che interessa tramite il servizio di *Resource Discovery*, per accedere ai dati effettivi presso il detentore degli archivi è necessario essere registrati e identificarsi con i consueti meccanismi di login. Per accedere ad archivi distribuiti, come nel nostro caso, occorre dunque in linea di principio essere registrati con tutti i fornitori di dati ed effettuare i diversi login, con una sarabanda di username e password potenzialmente diversi fra di loro e comunque frustranti. È perciò in corso di studio un'iniziativa che permetta un'unica identificazione con il sistema centrale (ad esempio quello di ARIADNE) il quale si farà carico di negoziare l'accesso con i vari sistemi che lo compongono, sul modello ormai consolidato di eduroam, la rete internazionale per il roaming dell'accesso a internet che consente agli appartenenti ai sistemi associati - in pratica a tutti i ricercatori accademici e degli enti di ricerca - di connettersi a internet in qualunque sede affiliata si trovino. Il sistema allo studio per le scienze umane, compresa l'archeologia, prevede di accogliere anche ricercatori *homeless*, ormai sempre più frequenti nel sistema della ricerca per la vasta presenza di figure non strutturate nel mondo della ricerca accademica. Aspetti quali il controllo di qualità dei dati, soluzioni per l'accesso unificato come quella sopra descritta, politiche per la gestione dei registri e per la standardizzazione della documentazione sono affrontati in un altro progetto europeo, PARTHENOS (*Pooling Activities, Resources and Tools for Heritage e-Research Networking, Optimization and Synergies*, cf. PARTHENOS 2015), che li considera nel quadro più generale delle scienze umane.

Lo scenario che si profila da qui a dieci anni è quello di una comunità collaborativa di ricercatori che integra le indispensabili ricerche sul campo e in laboratorio con l'accesso a biblioteche digitali specializzate aperte, riutilizzando i dati prodotti da altri ricercatori prima di loro. I sistemi integrati di dati saranno quello che sono state fino ad oggi le biblioteche, fonte di scambio di conoscenza e informazioni e di accesso al sapere sviluppato da altri, realizzando così in modo sempre aggiornato l'indicazione fornita quasi 200 anni fa da Gerhard: «nell'archeologia [...] gli sforzi di uno solo non possono conseguire veri progressi senza l'aiuto continuato e reciproco di molti altri, che mantenendo fra loro le più attive comunicazioni, facciano copia gli uni agli altri de' loro lumi e cognizioni» (Gerhard 1829, 3).

8 Riconoscimento

Il presente lavoro è stato supportato dal progetto ARIADNE, finanziato dalla Commissione Europea (contratto 313193) nel contesto del VII Programma Quadro di Ricerca e Sviluppo, bando INFRA-2012-1.1.3. Le opinioni dell'autore non riflettono necessariamente quelle della Commissione Europea.

Bibliografia

- ARIADNE (2015). URL <http://www.ariadne-infrastructure.eu> (2015-08-31).
- CIDOC-CRM (2015). URL <http://www.cidoc-crm.org> (2015-08-31).
- Gerhard, Friedrich W.E. (1829). «Osservazioni preliminari». *Annali dell'Istituto di Corrispondenza Archeologica*, 1, 3.
- Meghini, Carlo et al. (2017). «ARIADNE. A Research Infrastructure for Archaeology». De Santo, Massimo; Niccolucci, Franco; Richards, Julian D. (eds.), «Special Issue on Digital Infrastructure for Cultural Heritage», special issue, *Journal on Computing and Cultural Heritage (JOCCH)*, 10(2), issue 3, August 2017.
- Niccolucci, Franco (2013). «Un'infrastruttura di ricerca per l'archeologia. Il progetto ARIADNE». *Digitalia*, 2, 154-61.
- Niccolucci, Franco (2015). «Digital Archaeological Archives of Europe. Opportunities and Challenges». Schut, Peter A.C.; Scharff, Djurra; de Wit, Leonard C. (eds.), *Setting the Agenda. Giving New Meaning to the European Archaeological Heritage*. Budapest: Archaeolingua, 111-18.
- Niccolucci, Franco; Richards, Julian D. (2013). «ARIADNE. Advanced Research Infrastructures For Archaeological Dataset Networking in Europe». *International Journal of Humanities and Arts Computing*, 7(1-2), 70-88.
- PARTHENOS (2015). URL <http://www.parthenos-project.eu> (2015-08-31).

BIA-Net: l'interoperabilità delle fonti giuridiche antiche attraverso il Semantic Web

Daria Spampinato

(Istituto di Scienze e Tecnologie della Cognizione, CNR Catania, Italia)

Abstract In Roman Law domain we have brought forward the Bibliotheca Juris Antiqui (BIA-Net) project that, moving from a standard domain, evolves according to Linked Data principles. BIA-Net system is a framework that makes the Roman Law scientific heritage accessible online. The data are encoded according to the DC, TEI and SKOS vocabularies. At this functionality designed for the specific needs of Roman Law scholars, with search functions, views and dedicated services, we have added the publication of BIA-Net in RDF OWL ontology to ensure interoperability with other Semantic Web resources. Notably, Roman Law sources, in Latin and Greek, are encoded, on the one hand, in TEI format – a format that enables high expressiveness of the texts – and, on the other hand, are re-engineered into a specific ontology reusing Ancient Laws' modelled metadata schema and domain ontologies.

Sommario 1 Metadati e ontologie per le scienze dell'antichità. – 2 Bia-Net – Bibliotheca Juris Antiqui. – 3 TEI e Semantic Web. – 4 Conclusioni e prospettive.

Keywords Legal resources. Ontology. Semantic Web. TEI. SKOS.

1 Metadati e ontologie per le scienze dell'antichità

La complessità del dominio delle Scienze dell'Antichità e più in generale del Cultural Heritage e delle Digital Humanities e la eterogeneità dei contenuti da trattare hanno portato alla proliferazione di schemi o modelli di metadati. Uno schema di metadati è un insieme di elementi con un preciso significato per descrivere elementi, oggetti e contenuti; uno schema definisce la semantica degli elementi e talvolta le regole per la specificazione dei valori (cf. Haslhofer, Klas 2010, 7-8).

Il fenomeno di digitalizzazione ha prodotto negli ultimi anni circa 105 standard per la conservazione, diffusione e gestione dei contenuti culturali, ognuno valido per applicazioni verticali e per tutte le diverse istituzioni culturali (musei, archivi, biblioteche). Questi standard possono essere sud-

divisi per dominio, comunità, funzione e scopo.¹ Tra essi ve ne sono alcuni più usati nel settore delle Digital Humanities e tra questi: DC,² FOAF,³ FRBR,⁴ MARC,⁵ OAI/ORE,⁶ RDF,⁷ SKOS,⁸ TEI,⁹ XML.¹⁰

Oltre agli schemi di metadati vi è un largo numero di differenti vocabolari (thesauri, liste controllate, ontologie) quali GeoNames, Pleiades, WordNet, DBpedia, VIAF, Thesaurus PICO.¹¹ L'aggregazione e la connessione sul web di dati codificati secondo diversi schemi di metadati e vocabolari e con infrastrutture eterogenee pone il problema della interoperabilità tecnologica e semantica: come cioè integrare dati eterogenei e come far comunicare differenti risorse digitali (bibliografiche, testo, immagini, ecc.) (cf. Peroni, Tomasi, Vitali 2013, 413).

Un altro problema che si presenta nei sistemi informativi sul web è l'identità. Ad esempio, supponiamo di volere cercare sul web informazioni sulla *Lex de imperio Vespasiani* utilizzando un motore di ricerca tradizio-

1 Cf. «Seeing Standards (CC-BY-SA)», URL <http://jennriley.com/metadatamap> (2017-11-13).

2 Dublin Core (<http://dublincore.org/documents/dces/>) è il più diffuso schema di metadati di base descrittivo di qualsiasi risorsa digitale.

3 Friend of a Friend (<http://www.foaf-project.org/>) è uno standard RDF per descrivere persone fisiche utenti del web che abbiano nome, indirizzo e-mail, pagine web e altri dati identificativi.

4 Functional Requirements for Bibliographic Records (<https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>) è un modello concettuale bibliografico sviluppato dalla International Federation of Library Associations and Institutions (IFLA).

5 Machine Readable Cataloging (<http://www.loc.gov/marc/>) è uno standard ISO per la rappresentazione dell'informazione bibliografica.

6 Open Archives Initiative Object Reuse and Exchange (<http://www.openarchives.org/ore/>) definisce formati per la descrizione e lo scambio di aggregazione di risorse digitali sul web.

7 Resource Description Framework (<http://www.w3.org/TR/rdf-primer/>) è lo standard w3C di base del Semanti Web.

8 Simple Knowledge Organization System (SKOS) è un vocabolario RDF - W3C Recommendation utilizzato per rappresentare sistemi strutturati di organizzazione della conoscenza, *de facto* standard per il semantic web per tale scopo. Le specifiche di SKOS sono disponibili all'indirizzo <http://www.w3.org/TR/skos-reference/>.

9 Text Encoding Initiative (<http://www.tei-c.org>) è uno schema di metadati per la codifica di documenti testuali. Costituisce uno standard *de facto* della comunità delle Digital Humanities per la codifica dei testi.

10 eXtensible Markup Language (<http://www.w3.org/XML/>) è un meta-linguaggio per definire linguaggi di markup per specifici scopi.

11 URL <http://geonames.org>; <http://pleiades.stoa.org/>; <https://wordnet.princeton.edu/>; <http://dbpedia.org>; The Virtual International Authority File disponibile all'indirizzo <http://viaf.org>; http://www.culturaitalia.it/pico/thesaurus/4.3/thesaurus_4.3.0.skos.xml.

nale; digitando 'vespasiano' si ricavano diverse informazioni su entità assolutamente differenti (il teatro, l'imperatore, l'orinatoio e la lex): entità da disambiguare per capire l'identità a cui vogliamo riferirci. Un altro aspetto dell'identità delle risorse sul web riguarda l'univocità di un determinato elemento sul web, cioè il fatto che un'unica entità (ad esempio la *Lex de imperio Vespasiani*) potrebbe essere presente sul web con diverse istanze.

I problemi di interoperabilità e di identità che abbiamo sommariamente indicato possono essere affrontati utilizzando in modo semplice ed efficace le tecnologie del Semantic Web nella prospettiva Linked Data.

Il Semantic Web consiste nel creare i collegamenti tra dati sul web, in modo che una persona o un automa possa esplorare il web dei dati. Esso viene comunemente rappresentato a livelli: si fonda sugli URI,¹² gli standard aperti del W3C RDF, SPARQL¹³ e OWL¹⁴ e le tecnologie semantiche. Il modello di base di rappresentazione dei dati del Semantic Web è lo standard RDF, che descrive la semantica dell'informazione in modo da renderla accessibile in maniera automatica; ha come struttura di base la tripla ricorsiva soggetto-predicato-oggetto che, collegata con altre triple, forma un grafo RDF rappresentante un'ontologia.¹⁵

12 L'Uniform Resource Identifier (URI) è un sequenza di caratteri che identifica univocamente una risorsa in Internet.

13 SPARQL (acronimo di SPARQL Protocol And RDF Query Language) è un linguaggio di interrogazione per dati rappresentati in RDF. Le specifiche di SPARQL sono disponibili all'indirizzo <http://www.w3.org/TR/sparql11-query/>.

14 Le specifiche di Web Ontology Language (OWL) sono disponibili all'indirizzo <http://www.w3.org/TR/owl2-overview/>.

15 Le ontologie (computazionali), chiamate anche vocabolari, sono sistemi formali il cui fine è rappresentare la conoscenza relativa a un determinato dominio. Le ontologie sono espresse in OWL, un'estensione di RDF, per consentire l'uso di ragionatori automatici che derivano inferenze logiche dalla struttura dei dati. Le triple di un'ontologia vengono interrogate con il linguaggio SPARQL. La capacità espressiva di OWL e di SPARQL permette di porre domande complesse a sorgenti di conoscenza eterogenee. Esempi di domande nel dominio del Diritto Romano e delle scienze dell'antichità sono: «quali articoli scientifici trattano del reato di 'ambitus' e citano Cicerone?» e «quali fonti non giuridiche trattano di 'ambitus' e a quali aree geografiche si riferiscono?».

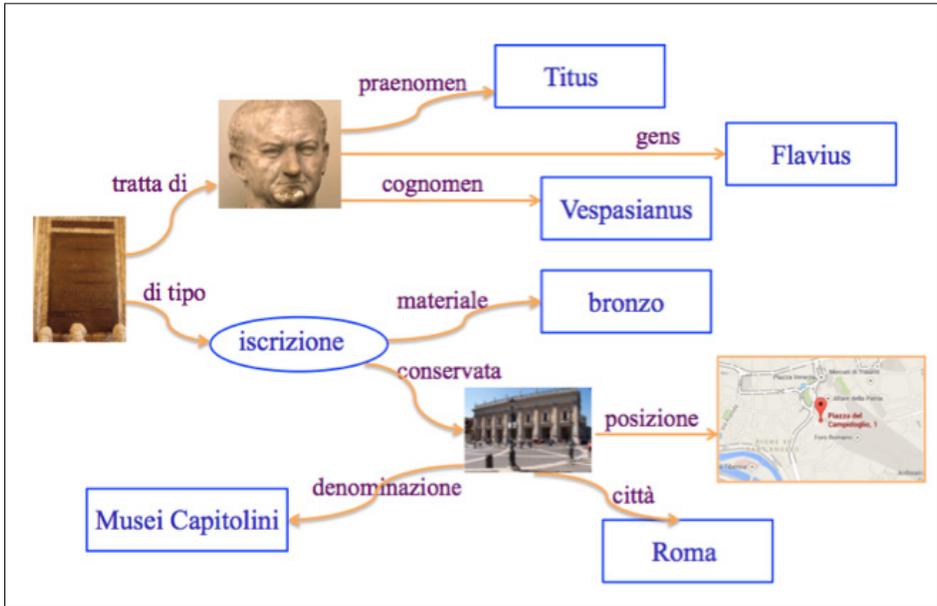


Figura 1. Grafo della Lex de imperio Vespasiani

In figura 2 si ha un esempio di grafo riguardante la *Lex de imperio Vespasiani*. Dalla navigazione attraverso i nodi del grafo si possono ricavare le informazioni che l'entità Imperatore Vespasiano ha *praenomen* Titus, *gens* Flavius e *cognomen* Vespasianus; la Lex è un'iscrizione di bronzo conservata ai Musei Capitolini di Roma, di cui si hanno le coordinate per la georeferenziazione. Infine è utile ricordare che per assicurare la prospettiva Linked Data, l'ideatore del web dei documenti e poi del web dei dati, Tim Berners-Lee (2009), ha definito i principi base cui i dati devono attenersi.¹⁶

16 I Linked Data sono dati in formato RDF che rispettano i seguenti principi:
 - usare URI come nomi per le 'cose';
 - usare URI HTTP in modo che sia possibile cercarli e risolverli;
 - quando qualcuno cerca una URI, fornire un'informazione utile, non un generico errore;
 - includere link ad altre URI, così da permettere a chi cerca di scoprire nuovi collegamenti.

2 BIA-Net – Bibliotheca Iuris Antiqui

Nel dominio del Diritto Romano e dei Diritti dell'Antichità è in corso di elaborazione da alcuni anni un progetto che muovendo da standard di dominio sta seguendo il modello Linked Data.

BIA-Net è un framework verticale che rende accessibile online in formas strutturato non proprietario il patrimonio scientifico romanistico, realizzando un sistema che vede come destinatari privilegiati i giusromanisti.

BIA è l'integrazione della conoscenza di tre archivi: *Fontes*, *Opera* e *Thesaurus*. Essi sono integrati in modo che la grande quantità di dati possa essere accessibile attraverso diverse parole chiavi. L'archivio *Fontes* contiene la maggior parte delle fonti giuridiche romane a testo pieno per un totale di circa quattordici milioni di caratteri. *Opera* è un repertorio bibliografico di cinquantamila record. Il *Thesaurus* è un archivio trasversale dei concetti dei diritti dell'antichità, il cui scopo è di espandere le ricerche delle informazioni degli altri archivi, utilizzando le relazioni concettuali (di preferenza, gerarchiche e associative) del dominio di conoscenza presenti nel tesoro. Le connessioni tra le informazioni contenute negli archivi sono fornite dai collegamenti concettuali presenti nei record di *Opera*, il cui ruolo è quello di collegare gli altri due archivi.

I tre archivi hanno differenti strutture interne dovute alla natura delle informazioni contenute, quindi per rappresentarli sono stati utilizzati specifici schemi di metadati. In particolare i record di *Opera* sono stati strutturati utilizzando il Dublin Core; i testi di *Fontes* sono stati codificati secondo le linee guida del TEI; il *Thesaurus* è espresso in SKOS.

Il framework Bia-Net è un sistema di *information retrieval* dedicato agli studiosi del diritto romano; offre funzioni di ricerca, visualizzazioni e servizi dedicati. Il sistema contiene tre distinte maschere di ricerca, una per ciascuno dei tre archivi integrati (*Fontes*, *Opera*, *Thesaurus*) e meccanismi di manutenzione e aggiornamento (ad esempio, un editor per inserire e modificare le schede bibliografiche di *Opera*). Nella maschera di ricerca *Fontes* si può cercare nei campi 'Luogo fonte' (o *Locus*), 'Rubrica', 'Inscriptio', 'Text', 'Subscriptio' e anche in un campo a testo libero, come si vede nella schermata in figura 3. Attraverso la maschera di ricerca *Opera* si possono recuperare le schede bibliografiche secondo i tipici elementi quali autore, titolo, pubblicazione, e secondo i codici di classificazione e i termini controllati del *Thesaurus* e le fonti maggiormente citate (la maggior parte delle quali è contenuta in *Fontes*).



Figura 2. Maschera di ricerca *Fontes* del framework BIA-Net

Inoltre dalla schermata di visualizzazione del record (fig. 4) si può navigare verso gli altri archivi, utilizzando i collegamenti concettuali presenti nei valori dei vari campi.

Title	«Sponsio» e «stipulatio». Divagazioni intorno alla storia del contratto, dell'obbligo, delle garanzie personali
Author	BIONDI B. Q
Publication	BIDR. 65 Q
References	BIDR. 65, 105-129 (1962)
Year	1962 Q
Classing	3.4.2.1.1 Q
Checked Terms	Sponsio Q Q Q
Publication type	articolo

[Q Search in Fontes](#) - [Q Search in Opera](#) - [Q Search in Thesaurus](#)

Figura 3. Visualizzazione di un record bibliografico in *Opera*

In aggiunta, per essere in grado di utilizzare tutte le potenzialità della rete e l'interoperabilità con altre ontologie presenti nel Semantic Web al fine dell'arricchimento dei dati, si sta anche sviluppando un'altra funzionalità degli archivi, riorganizzando i dati di BIA in un dataset nella prospettiva del Linked Data. Le due funzionalità coesisteranno perché ciascuna risponde a esigenze differenti. La figura 5 mostra la doppia funzionalità degli archivi: da un lato essi sono parte integrale del framework; dall'altra sono organizzati come Linked Open Data.

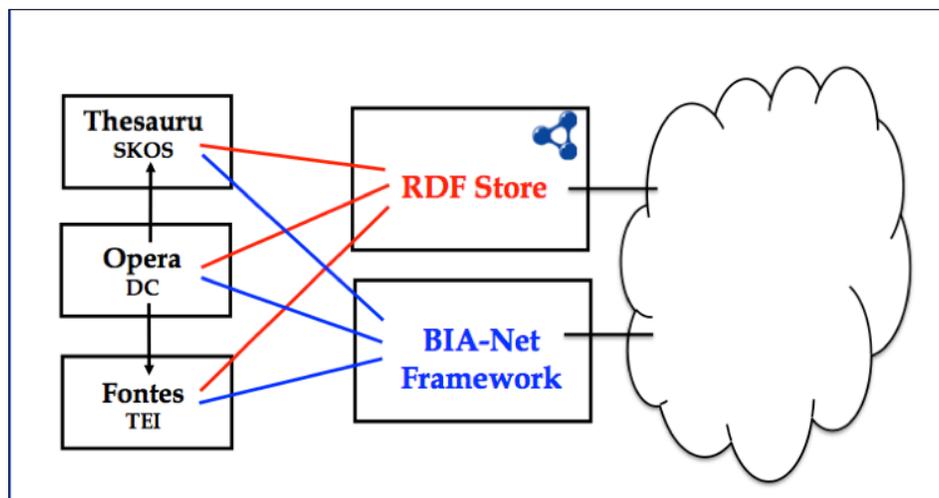


Figura 4. Le due funzionalità degli archivi di BIA

A seconda dell'archivio sono state adottate soluzioni differenti per la reingegnerizzazione dei dati. I dati del *Thesaurus* e di *Opera* possono essere con pochi passaggi memorizzati in un triple store. Nelle figg. 5 e 6 sono rappresentate porzioni di grafi RFD del *Thesaurus* e di *Opera*.

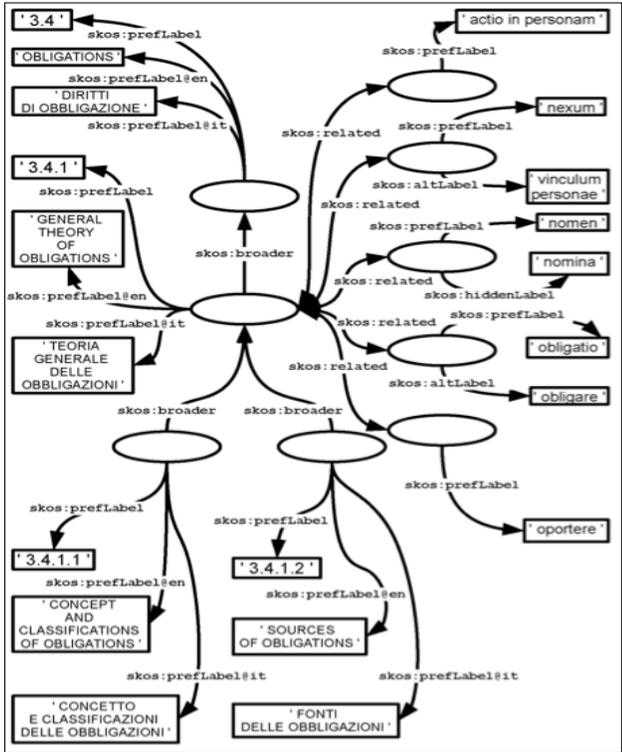


Figura 5. Porzioni di grafi RFD del *Thesaurus*

Il *Thesaurus* di BIA è stato convertito in RDF usando qualche pattern dello schema di metadati SKOS (cf. Spampinato, Zangara 2013). SKOS, implementato con RDF/XML, è una buona rappresentazione e gestione di sistemi strutturati di organizzazione della conoscenza, quale i tesauri, secondo il paradigma dei Linked Open Data. I dati di *Opera* sono in Dublin Core espresso usando il modello sintattico RDF.¹⁷

¹⁷ Una dettagliata descrizione dei processi di conversione dei dati di *Fontes* in TEI e dei record di *Opera* in DC/XML è presente in (Spampinato 2012a) e (Spampinato 2012b).

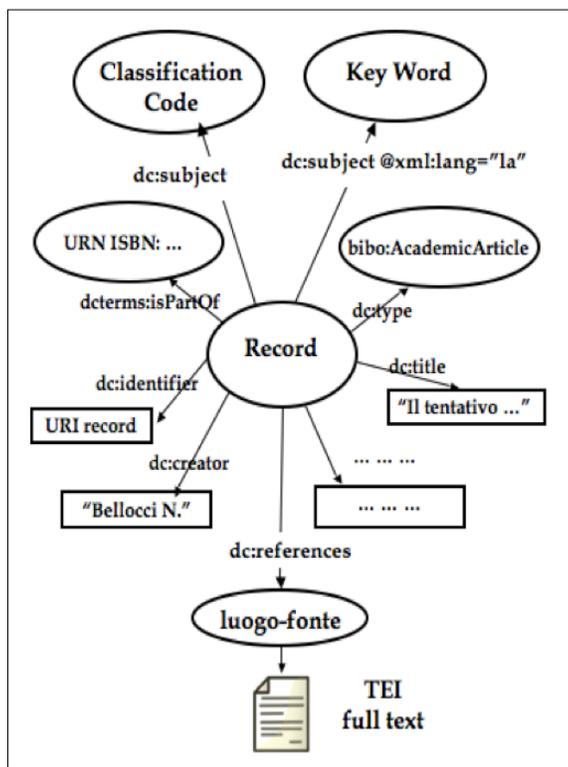


Figura 6. Porzioni di grafi RFD di *Opera*

3 TEI e Semantic Web

Per quanto concerne i dati di *Fontes* bisogna fare qualche riflessione aggiuntiva. Questo archivio di testi del diritto romano in latino e in greco è in formato TEI (fig. 7). Tale formato permette una espressività delle fonti che nessun altro modello catturerebbe, codificando i testi secondo la struttura logica interna e fornendo anche il contesto con altre informazioni semantiche, paratestuali e critiche. Quindi si è scelto di mantenere la codifica dei dati di *Fontes* in TEI e di riorganizzarli anche in un'ontologia.

```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
...
<body>
<text>
...
<div1 xml:id="D01" n="D.1" type="Libro">
<head>Liber primus</head>
...
<div2 xml:id="D0102" n="D.1.2" type="Titolo">
<div3 xml:id="D0102R" n="D.1.2r" type="Rubrica">
<argument>
<p>DE ORIGINE IURIS ET OMNIUM MAGISTRATUUM ET SUCCESSIONE PRUDENTIUM</p>
</argument>
</div3>
...
<div3 xml:id="D010202" n="D.1.2.2" type="Frammento">
<div4 xml:id="D010202I" n="D.1.2.2" type="Inscriptio">
<p rend="italic"><name rend="sc">Pomponius</name>libro singulari enchiridii</p>
</div4>
...
<div4 xml:id="D01020237" n="D.1.2.2.37" type="Paragrafo">
<p>Fuit post eos maxime scientiæ <name rend="sc">Sempronius</name>, quem populus
Romanus <foreign xml:lang="grc" rend="italic">ἡμῶν</foreign>
appellavit, nec quisquam ante hunc aut post hunc hoc nomine cognominatus est.
<name rend="sc">Gaius Scipio Nasica</name>, qui optimus a senatu appellatus est: cui etiam publice
domus in sacra via data est, quo facilius consuli posset. deinde <name rend="sc">Quintus Mucius</name>,
qui ad Carthaginienses missus legatus, cum essent duæ tesserae positæ una pacis altera belli,
arbitrio sibi dato, utram vellet referret Roman, utramque sustulit et ait Carthaginienses petere debere,
utram mallent accipere.</p>
</div4>
...
</div3>
</div2>
...
</div1>
...
</text>
</body>
</TEI>

```

Figura 7. Brano dei Digesta codificato in TEI/XML

All'interno della comunità TEI vi sono diverse strategie di integrazione tra TEI e altri modelli semantici (CIDOC-CRM,¹⁸ FRFBoc¹⁹ e RDF) per l'interscambio dei dati²⁰ e per accrescere le potenzialità delle edizioni digitali nell'ottica dei Linked Data. Questo perché il modello orientato al documento del TEI e i modelli semantici sono concettualmente differenti, ma entrambi possono essere combinati per essere usati in applicazioni delle Digital Humanities e nel Cultural Heritage.

Le modalità di interazione tra TEI e le ontologie si possono schematizzare in: tecniche *embedding*, che consistono nell'incorporare le informazioni nel documento codificato in TEI; tecniche *standoff*, che permettono di memorizzare le annotazioni in un file separato con riferimenti alle parti

18 CIDOC-Conceptual Reference Model (<http://www.cidoc-crm.org>) definisce concetti e relazioni che descrivono elementi del Cultural Heritage.

19 FRBR-object oriented è un'armonizzazione tra FRBR e CIDOC-CRM per rappresentare informazioni museali e bibliografiche.

20 Le modalità di collegamento tra i due modelli sono state analizzate sotto differenti punti di vista in Eire 2015; Ciotti, Tomasi 2016; Eide, Ore 2009.

del testo TEI. Entrambe presentano pro e contro e la scelta deve essere guidata dalle funzionalità richieste. Le tecniche possono anche essere combinate per sfruttare i vantaggi di entrambe.

Esistono alcuni progetti²¹ che coniugano la codifica in TEI per i testi dei documenti e dei manoscritti con modelli del Semantic Web; ognuno di questi progetti utilizza una strategia diversa di interazione tra TEI e RDF, consentendo così diversi livelli di interpretazione e utilizzo del corpus.

Per quanto riguarda BIA-Net inizialmente viene estratta la conoscenza da BIA *Fontes* a partire dai luoghi fonte. Vengono, cioè, individuate le entità che definiscono i luoghi fonte. Un luogo fonte identifica univocamente ogni elemento base di un testo giuridico: ad esempio D.20.1.3pr indica il *principium* del terzo frammento del primo titolo del ventesimo libro del Digesto. Come si può notare anche in figura 7, le fonti sono state codificate opportunamente in TEI, in modo che ogni brano sia individuato da un tag <div> numerato avente un attributo n che riporta il luogo fonte e un attributo xml:id unico che ripropone il luogo fonte senza ambiguità.

Successivamente si opererà una estrazione di tutta la conoscenza contenuta in *Fontes* per creare un'ontologia specifica per i diritti dell'antichità.

4 Conclusioni e prospettive

Con BIA-Net si intendono soddisfare due diverse esigenze: da un lato fornire alla comunità dei giusromanisti e degli studiosi dei diritti dell'antichità uno strumento completo che soddisfi le esigenze specifiche della comunità con funzioni di ricerca, visualizzazioni e servizi dedicati; dall'altro assicurare l'interoperabilità con altre risorse presenti nel Semantic Web mediante la pubblicazione di BIA-Net secondo il paradigma dei Linked Open Data.

Per fare ciò tutti i dati presenti negli archivi di BIA devono essere convertiti in triple RDF oppure, come nel caso di *Fontes*, collegati con un'opportuna ontologia.

L'ontologia di tutto il patrimonio informativo di BIA viene modellata riusando schemi di metadati consolidati, come FOAF e BIBO.²² Le identità delle persone, luoghi, concetti giuridici estratti dal testo vengono risolte secondo ontologie e risorse generaliste e di dominio (DBpedia, WordNet, Perseus Digital Library,²³ Pleiades, GeoNames e lo stesso *Thesaurus* BIA).

In un'ottica di interoperabilità con altre risorse sul web e per rendere

21 Henry III Fine Roll (cf. Ciula, Spence, Vieira 2008), Sharing Ancient Wisdoms (SAWS) (cf. Hedges et al. 2012) e CLAROS <http://www.clarosnet.org>.

22 Bibliographic Ontology (BIBO) è disponibile all'indirizzo <http://bibliontology.com/>.

23 URL <http://www.perseus.tufts.edu/>.

i contenuti di BIA-Net accessibili a quanti più studiosi possibile, una strategia consiste nel collegamento con Europeana²⁴ attraverso la mappatura dello schema di metadati di BIA nello schema EDM²⁵ di Europeana. Essa, infatti, funge da raccordo tra i provider delle istituzioni culturali (musei, biblioteche, archivi) per fornire un unico accesso ai contenuti culturali europei; e contemporaneamente mira a pubblicare i dati in modalità Linked Data.²⁶ Oltre alle informazioni presenti internamente, Europeana mantiene i riferimenti diretti ai dati presenti nelle istituzioni culturali, in modo che di ogni contenuto culturale siano presenti diverse interpretazioni.

Bibliografia

- Barabucci, Gioele et al. (2012). «Embedding Semantic Annotations Within Texts. The Fretta Approach». Ossowski, Sascha; Lecca, Paola (eds.), *Proceedings of the 27th Annual ACM Symposium on Applied Computing = SAC '12. ACM Symposium on Applied Computing* (Riva, Italy, 26th-30th March 2012). New York: ACM, 658-63.
- Berners-Lee, Tim (2009). «Linked Data - Design Issues» [online]. *World Wide Web Consortium (W3C)*. URL <http://www.w3.org/DesignIssues/LinkedData.html> (2017-10-16).
- Ciotti, Fabio; Tomasi, Francesca (2016) «Formal Ontologies, Linked Data, and TEI Semantics» [online]. *Journal of the Text Encoding Initiative*, 9. URL <http://jtei.revues.org/1480> (2017-11-10).
- Ciula, Arianna; Spence, Paul; Vieira, José Miguel (2008). «Expressing Complex Associations in Medieval Historical Documents. The Henry III Fine Rolls Project». *Literary and Linguistic Computing*, 23(3), 311-25.
- Eide, Øyvind; Ore, Christian-Emil (2009). «TEI and Cultural Heritage Ontologies. Exchange of Information?» [online]. *Literary & Linguistic Computing*, 24(2), 161-72. DOI 10.1093/lc/fqp010.
- Eide, Øyvind (2015). «Ontologies, Data Modeling, and TEI» [online]. *Journal of the Text Encoding Initiative*, 8. URL <http://jtei.revues.org/1191> (2017-10-16).
- Haslhofer, Bernhard; Klas, Wolfgang (2010). «A Survey of Techniques for Achieving Metadata Interoperability» [online]. *ACM Computing Surveys*, 42(29), 1-37, article 7. DOI 10.1145/1667062.1667064.

24 URL <http://www.europeana.eu>.

25 La documentazione di Europeana Data Model (EDM) è disponibile all'indirizzo <http://pro.europeana.eu/page/edm-documentation>.

26 La descrizione del progetto e tutta la documentazione relativa sono disponibili all'indirizzo <http://labs.europeana.eu/api/linked-open-data/introduction/>.

- Hedges, Mark et al. (2012). «Exploring Manuscripts. Sharing Ancient Wisdoms across the Semantic Web» [online]. *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics* (Craiova, Romania, 13th-15th June 2012). *WIMS'12*, 1-12, article 44. New York: ACM. DOI 10.1145/2254129.2254184.
- Peroni, Silvio; Tomasi, Francesca; Vitali, Fabio (2013). «The Aggregation of Heterogeneous Metadata in Web-based Cultural Heritage Collections. A Case Study» [online]. *International Journal of Web Engineering and Technologies*, 8(4), 412-32. DOI 10.1504/IJWET.2013.059107.
- Spampinato, Daria (2012a). «La codifica delle fonti giuridiche romane in TEI/XML e dei dati bibliografici in DC/XML e il progetto BIA-Net». *Minima Epigraphica et Papyrologica*, XII-XV(14-17), 293-326.
- Spampinato, Daria (2012b). «L'uso degli standard XML per la gestione in rete dei documenti giuridici romani». Palazzolo, Nicola; Istituto italiano di scienze umano (eds.), *Diritto romano e scienze antichistiche nell'era digitale = Convegno di studio* (Firenze, 12-13 settembre 2011). Torino: Giappichelli, 169-88. *Collectanea graeco-romana* 10.
- Spampinato, Daria; Zangara, Ignazio (2013). «Classical Antiquity and Semantic Content Management on Linked Open Data» [online]. Tomasi, Francesca; Vitali, Fabio (eds.), *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environments. Metadata, Vocabularies and Techniques in the Digital Humanities (DH-Case 2013)*. New York: ACM. DOI 10.1145/2517978.2517992.

Il progetto EAGLE

Europeana network of Ancient Greek and Latin Epigraphy e le sue molteplici sfide

Silvia Orlandi

(Università degli Studi di Roma "La Sapienza", Italia)

Abstract EAGLE (Electronic Archive of Greek and Latin Epigraphy) was born in 2003 as a federation of four epigraphic digital archives (Epigraphic Database Bari-EDB, Epigraphische Datenbank Heidelberg-EDH, Epigraphic Database Roma-EDR, and Hispania Epigraphica Online). In 2013 it became a more complex and comprehensive European project, co-funded by the European Commission for 36 months, in the framework of the ICT-PSP program, whose aim is to aggregate epigraphic contents provided by different databases (see the complete list on the website <http://www.eagle-network.eu>) and make them searchable through a single portal. These contents, harmonized and "disambiguated" in order to give a permanent identifier to records coming from different archives but related to a single object, are then provided to Europeana, the portal of European cultural heritage, to make them accessible not only to scholars, but also to the broad public (that's why the acronym is now expanded Europeana network of Ancient Greek and Latin Epigraphy).

Keywords Greek epigraphy. Latin epigraphy. Digital humanities. Europeana. Translations.

In occasione del primo convegno organizzato dall'Associazione per l'Informatica Umanistica e la Cultura Digitale, Antonio Felle fu invitato a presentare due importanti realtà nel campo dell'informatica applicata all'epigrafia: EAGLE e EpiDoc (Felle 2012). Allora, nel 2011, EAGLE era una federazione di quattro banche dati - Epigraphic Database Bari (EDB), Epigraphic Database Heidelberg (EDH), Epigraphic Database Roma (EDR) e Hispania Epigraphica Online (HEp) - che, riconoscendosi in un comune modo di intendere lo studio delle epigrafi, avevano deciso di consorziarsi per rendere i propri dati liberamente disponibili e consultabili attraverso un unico portale (cf. la presentazione e i documenti raccolti in http://www.eagle-eagle.it/Italiano/present_it.htm). Contestualmente, per evitare il più possibile sovrapposizioni e cercare anche di rendere più rapido il processo di digitalizzazione, i membri della federazione avevano concordato una divisione degli ambiti di competenza di ciascuna banca dati, per cui EDB si sarebbe occupata delle iscrizioni di Roma di committenza cristiana, EDR delle iscrizioni di Roma (tranne quelle cristiane) e dell'Italia antica, nonché delle province di Sicilia e Sardinia, EDH delle epigrafi delle province, tranne Sicilia e Sardinia, competenza, come abbiamo detto, di

Antichistica 14

DOI 10.14277/6969-182-9/ANT-14-4 | Submission 2017-08-14

ISBN [ebook] 978-88-6969-182-9 | ISBN [print] 978-88-6969-183-6

© 2017 |  Creative Commons Attribution 4.0 International Public License

EDR, e le province iberiche, competenza di Hispania Epigraphica. Sempre su questa base, lo stato dei lavori della federazione EAGLE, aggiornato a giugno 2012, è stato poi sinteticamente presentato in un flyer distribuito in occasione del XIII Congresso Internazionale di Epigrafia Greca e Latina, tenutosi a Berlino nell'agosto 2012, pubblicato anche in forma di breve articolo (Pancierera 2013).

La relazione di Antonio Felle, pur sottolineando la validità di un'iniziativa destinata a dare accesso alle epigrafi nella loro vera natura non di puri testi, ma di monumenti iscritti, in cui le componenti non verbali (scrittura, supporto, contesto topografico) hanno un'importanza pari al contenuto nel veicolare il messaggio contenuto nell'iscrizione, non nascondeva le ombre e i limiti tecnici del progetto stesso, che, precludendo alcune possibilità di ricerca, ne riducevano in parte l'efficacia (per una serie di critiche e osservazioni alle banche dati esistenti nell'ambito dell'epigrafia digitale, cf. anche Felle 2014 e Hainzmann 2014). Auspicava, dunque, per il futuro, la disponibilità di maggiori risorse finanziarie che consentissero di migliorare il progetto EAGLE dal punto di vista quantitativo e qualitativo e concludeva con un rapido, quasi scaramantico accenno alla presentazione di una domanda per un finanziamento europeo (Felle 2012, 123).

L'opportunità di trovare uno sponsor nella Commissione Europea è stata effettivamente individuata nella collaborazione, in qualità di content provider, al progetto Europeana (<http://www.europeana.eu/portal/>), il grande portale europeo della cultura che ormai da diversi anni raccoglie e mette a disposizione degli utenti oggetti digitali di varia natura (immagini, testi, video, brani musicali, ecc.) provenienti da biblioteche, musei e archivi, pubblici e privati, dei diversi paesi dell'Unione Europea che hanno in corso la digitalizzazione delle loro collezioni. Tra i milioni di *items* presenti in questo enorme archivio virtuale, le epigrafi, in particolare quelle antiche, erano presenti solo in minima parte, quasi mai identificate con un riferimento bibliografico, spesso scelte in quanto 'opere e oggetti d'arte' (are, rilievi, sculture, ecc.) dalle istituzioni museali che contribuiscono al portale come fornitori di contenuti, mai come oggetti identificabili per la loro particolare natura di documenti utili alla nostra conoscenza del mondo antico. Di qui l'idea di presentare il progetto EAGLE come 'gap filler' di Europeana, proponendosi come contributori in grado di colmare una lacuna oggettivamente presente nel portale europeo - dove l'antichità in genere è largamente sottorappresentata - aggregando una serie di contenuti presenti in varie banche dati epigrafiche e rendendoli disponibili attraverso un unico portale. Una prima domanda di finanziamento, presentata nel 2011 nell'ambito del bando *Aggregating Content for Europeana* del programma ICT PSP della Commissione Europea, non è andata a buon fine, ma ha ricevuto un punteggio tale - a solo mezzo punto dai progetti finanziati - che ci ha indotto a non desistere e a ritentare l'impresa l'anno successivo. Così, mettendo a frutto le osservazioni che avevano deter-

minato la valutazione non del tutto positiva del progetto precedente, e con la fondamentale collaborazione di alcune persone particolarmente esperte nella preparazione di questo tipo di domande - Franco Zoppi e Vittore Casarosa dell'ISTI-CNR di Pisa, Claudio Prandoni e Antonella Fresa della Promoter S.r.l., società cui è stato affidato il management del progetto, Raffaella Santucci e Luca Giberti della Sapienza Università di Roma - nel 2012 abbiamo ripresentato, nell'ambito dello stesso programma di finanziamento, il progetto EAGLE, divenuto nel frattempo l'acronimo di Europeana network of Ancient Greek and Latin Epigraphy. A proporre il progetto è stato un consorzio di 19 partner appartenenti a 12 paesi diversi: Austria, Belgio, Cipro, Croazia, Francia, Germania, Italia, Regno Unito, Romania, Slovenia, Spagna e Ungheria (per una lista completa dei partner, cf. <http://www.eagle-network.eu/about/partners/> e Giberti, Orlandi, Santucci 2014). Tra i partner, si distinguono enti di natura diversa - università, accademie e centri di ricerca, fondazioni e piccole e medie imprese - che svolgono ruoli diversi all'interno del progetto. Tra i veri e propri *content providers*, i fornitori, cioè, dei contenuti digitali destinati a Europeana, figurano i quattro membri 'storici' della federazione EAGLE, cioè le istituzioni universitarie presso cui hanno sede EDB (Università degli Studi Aldo Moro di Bari), EDH (Ruprecht-Karls Universität Heidelberg), EDR (Università degli Studi di Roma "La Sapienza") e HEP (Universidad de Alcalá). Un'importante novità è rappresentata dall'ingresso nel consorzio dei colleghi dell'Università Juraj Dobrila di Pula, in Croazia, dell'Istituto di Archeologia dell'Accademia Slovena di Scienze e Arti di Ljubljana e delle università Eötvös Loránd di Budapest e Babeş-Bolyai di Cluj Napoca. I primi due contribuiscono con materiali di prima mano relativi alle iscrizioni di Pola, Parentium, Nesactium ed Emona, che confluiscono in EDR in quanto parte dell'Italia antica, mentre gli altri hanno messo a disposizione i loro archivi fotografici per alcuni importanti centri della Pannonia e della Dacia romana, che vanno ad arricchire la Epigraphische Datenbank di Heidelberg. Ma uno dei 'punti di forza' di EAGLE nella sua nuova veste è stata l'adesione al consorzio di un numero significativo - e crescente, come vedremo meglio in seguito - di altri progetti di digitalizzazione di iscrizioni e di materiali a esse relativi, che si sono affiancati ai partner della federazione originaria, andando così ad ampliare notevolmente la quantità di *digital items* che saranno disponibili alla fine del lavoro. Partecipano, infatti, sin dall'inizio:

- il progetto The Last Statues of Antiquity dell'Università di Oxford, pubblicato nel maggio 2012, che comprende oltre 2.600 testimonianze relative all'uso della statuaria tra la fine del III e l'inizio del VI secolo, moltissime delle quali rappresentate da basi con le relative iscrizioni (<http://laststatues.classics.ox.ac.uk/>);

- la banca dati Ubi erat lupa (<http://www.ubi-erat-lupa.org/about.php>), con sede, all'epoca, presso l'Università di Salisburgo, che raccoglie migliaia di foto di monumenti iscritti pertinenti prevalentemente (ma non esclusivamente) alle province nordorientali dell'Impero, frutto soprattutto di sistematiche campagne fotografiche condotte dall'instancabile Ortoff Harl e da sua moglie Friederike;
- la nuova versione dello storico progetto PETRAE che, grazie all'impegno dei colleghi del CNRS e dell'Università di Bordeaux 3, ha reso disponibili online le iscrizioni latine dell'Aquitania fin qui pubblicate (<http://petrae.huma-num.fr/index.php/fr/>);
- la preziosa raccolta di foto storiche delle iscrizioni della Tripolitania e dell'Etruria meridionale che fanno parte delle *digital collections* della British School at Rome (<http://www.bsrdigitalcollections.it/>);
- il progetto Archaia Kypriaki grammateia, che riguarda fonti antiche sull'isola di Cipro, messo a punto da un ente leader nel campo dell'archeologia digitale come il Cyprus Institute di Nicosia (<https://goo.gl/XzPEoy>);
- *last but not least*, e senza nulla togliere agli altri partner, l'immenso apporto dato a EAGLE dall'adesione di Arachne, la banca dati che sta procedendo non solo alla digitalizzazione dell'archivio fotografico dell'Istituto Archeologico Germanico nelle sue diverse sedi, ma anche alla scansione di centinaia di sillogi epigrafiche a stampa pubblicate tra il XVI e il XIX secolo e quindi libere da copyright, inclusi i volumi ottocenteschi del *Corpus Inscriptionum Latinarum* (<http://arachne.uni-koeln.de/drupal/>).

Rendere contemporaneamente accessibili le informazioni contenute in tutti questi archivi non è un'impresa facile, e fondamentale, dunque, è il ruolo svolto dai partner tecnologici del progetto, in primis l'Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" del CNR di Pisa, che vanta una pluriennale esperienza nel campo dei processi di aggregazione di materiali digitali eterogenei. Dello sviluppo del portale, invece, si occupano, in stretta collaborazione con il CNR e in accordo con le esigenze di ricerca individuate da tutti i partner, gli informatici delle società Promoter e Gogate, due delle imprese che rappresentano l'indispensabile componente 'non accademica' del consorzio. Non va dimenticato, inoltre, l'importante ruolo dei colleghi della Katholieke Universiteit di Leuven, che hanno messo a disposizione la loro esperienza, maturata nell'ambito del progetto Trismegistos (<http://www.trismegistos.org/>), per 'disambiguare' i contenuti assegnando un identificatore unico ai materiali relativi a una stessa iscrizione presenti in più banche dati diverse.

L'apporto di tante persone, paesi e istituzioni rende indubbiamente complessa la gestione dell'impresa, ma garantisce anche uno dei punti cardine del progetto, e cioè la qualità e l'affidabilità dei contenuti, che sono forniti idealmente da coloro che, per esperienza di studio e conoscenza, spesso diretta, dei materiali, hanno più competenza sul nucleo di *items* di cui curano la digitalizzazione. L'enorme diffusione della 'rete' cui abbiamo assistito negli ultimi anni, infatti, ha portato a un aumento esponenziale delle informazioni disponibili via internet, che non necessariamente facilita la diffusione e soprattutto il miglioramento delle nostre conoscenze, vista la quantità di dati imprecisi, obsoleti o addirittura errati che si trovano online. Posto che solo una buona base documentaria consente di raggiungere validi risultati scientifici, e che l'affidabilità dei dati dipende non tanto dalla tecnologia, quanto dal criterio con cui tali dati sono messi in rete, abbiamo pensato che fosse responsabilità di chi si occupa quotidianamente di ricerca in campo epigrafico mettere a disposizione della comunità scientifica contenuti di alta qualità, anche a costo di sacrificare la rapidità con cui raggiungeremo l'obiettivo della completa digitalizzazione del patrimonio epigrafico.

A questo proposito mi sia consentito di fare una piccola digressione dal generale al particolare, illustrando brevemente l'organizzazione del progetto EDR, che, grazie a un efficace sistema di gestione online dell'immissione e della revisione dei dati, ha portato alla sua massima applicazione questo principio, coinvolgendo nel processo di digitalizzazione delle iscrizioni dell'Italia antica centinaia di persone - studenti di tutti i livelli, collaboratori, docenti italiani e stranieri - che, per vari motivi, si trovano nelle migliori condizioni possibili per procedere a una schedatura che non sia mera immissione di dati, ma una vera e propria edizione critica degli oggetti iscritti. Così, ad esempio, i colleghi di Genova si sono occupati delle iscrizioni della Liguria parallelamente alla loro pubblicazione nella collana dei *Supplementa Italica*; il gruppo di Venezia lavora alle iscrizioni di Belluno, Treviso e altri centri antichi della regione che pure sono stati e continuano ad essere oggetto di pubblicazioni recenti; per quanto riguarda Roma, i *tituli ad ludos pertinentes* sono affidati alle cure di un riconosciuto esperto in materia come Gian Luca Gregori, mentre delle *carminum reliquiae* si è occupato, tra il 2013 e il 2016, con risultati molto interessanti, il collega Alfredo Morelli, e così via (i responsabili della schedatura dei vari centri o delle varie tipologie di iscrizioni sono indicati nella pagina Stato dei lavori del sito <http://www.edr-edr.it>). Naturalmente, per quanto competente, nessuno è infallibile, e sviste, errori e omissioni sono sempre possibili. Porsi come obiettivo primario la qualità dei contenuti significa, quindi, prevedere anche la possibilità di una loro costante migliorabilità. È per questo che i responsabili tecnici di EDR, Lanfranco Fabriani e Silvia Evangelisti, hanno messo a punto un prezioso sistema di segnalazione di aggiunte e correzioni da parte degli utenti, che vengono

ricevute e valutate da un responsabile che sceglie se accoglierle o meno, e che rappresentano anche un interessante indicatore della frequenza con cui la banca dati viene consultata, e della tipologia e dell'origine geografica dei suoi utenti.

Lo sforzo di assicurare la qualità e l'affidabilità dei contenuti che caratterizza non solo EDR ma più in generale tutto il progetto EAGLE trova espressione anche nell'idea di fondo da cui il progetto è nato e che l'ampliarsi e il diversificarsi del consorzio non ha sostanzialmente modificato. Se, cioè, per essere correttamente intesa e utilizzata un'epigrafe va non solo analizzata nella sua componente testuale, ma anche contestualizzata e considerata nelle sue caratteristiche materiali, nello stesso modo essa andrà archiviata e presentata, anche in formato digitale. È tenendo conto di questo modello concettuale che è stato disegnato e strutturato il portale che dà accesso ai dati che confluiscono nel progetto (consultabile alla pagina <http://www.eagle-network.eu/search-inscriptions/>) e che prevede la possibilità di corredare le ricerche sui testi con una serie significativa di informazioni sulla localizzazione - antica e moderna - dell'iscrizione e sul suo supporto (tipologia, materiale, tecnica scrittoria), nonché sulle immagini (foto, disegni, ricostruzioni) che ne illustrano le caratteristiche, ferma restando la possibilità di risalire comunque al record completo che continua a essere presente nella banca dati d'origine.

Per rendere realmente disponibile questa massa di informazioni è stato necessario un enorme sforzo preliminare di armonizzazione dei contenuti e dei vocabolari utilizzati dalle varie banche dati, che ha occupato praticamente tutto il primo anno del progetto e che ha portato alla definizione di un comune modello di metadati, funzionale sia alla reperibilità delle informazioni attraverso il portale, sia alla loro trasmissione a Europeana, fine ultimo dell'intero percorso. Anche in questo caso, fondamentale è stata l'esperienza maturata negli ultimi anni da EpiDoc, un sistema di codifica dei materiali epigrafici che si sta affermando come standard sul piano internazionale e che i membri del consorzio EAGLE hanno scelto come base di riferimento per ridurre ad un unico comune denominatore dati che nascono e continuano ad essere immessi con sistemi estremamente eterogenei (l'intero processo è illustrato da Casarosa et al. 2014).

L'accesso contemporaneo alle molteplici iniziative digitali che nel corso degli ultimi anni sono nate nel campo dell'epigrafia, del resto, è una delle prime risposte alla domanda 'What do we need?' che un autentico pioniere nel campo dell'informatica applicata all'epigrafia come Silvio Panciera si è posto nei suoi recenti interventi sull'argomento (Panciera 2006, 2012). Che questo sforzo di armonizzazione venga incontro a un'esigenza fortemente sentita dagli 'antichisti digitali' è dimostrato, a mio avviso, anche dal fatto che nel corso del tempo il progetto EAGLE si è di fatto trasformato, come il primo congresso internazionale di Parigi ha dimostrato, in una sorta di 'piattaforma comune' in cui hanno interesse a

far confluire i propri materiali tutti quei progetti che, sia pure per ambiti geografici e cronologici diversi, si riconoscono in uno stesso modo di concepire l'epigrafia e aderiscono allo stesso spirito di libera condivisione dei risultati (gli Atti - *Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference*, a cura di Casarosa et al. - sono disponibili online ai seguenti link: <http://www.eagle-network.eu/eagle-first-international-conference-proceedings-now-online/> o <http://digilab-epub.uniroma1.it/index.php/Antichistica/article/view/220/209>).

È così che già nell'arco dei primi mesi di lavoro il consorzio EAGLE si è allargato alla partecipazione di vari altri enti e istituzioni che collaborano al progetto a vario titolo e con varie forme di accordo, fornendo contenuti dei loro archivi digitali (come nel caso, ad esempio, della Pontificia Commissione di Archeologia Sacra), condividendo l'esperienza sviluppata in altre discipline dell'antichistica, come la filologia, l'archeologia e la storia antica, ma soprattutto rendendo accessibili e interscambiabili le informazioni presenti in progetti affini, come quello relativo alle iscrizioni latine dell'Ashmolean Museum di Oxford (<http://www.ashmolean.org/ashwpress/latininscriptions/>) o l'Ancient Graffiti Project (<http://ancientgraffiti.wlu.edu/>) diretto da Rebecca Benefiel che, non a caso, collabora contemporaneamente con EDR per i graffiti dell'area vesuviana.

In tutti questi progetti è fondamentale il ruolo svolto dalle immagini, che veicolano, insieme ai metadati che le accompagnano, una serie di informazioni indispensabili alla completa e corretta comprensione del messaggio contenuto nell'iscrizione (dimensione dei caratteri, disposizione del testo nel campo epigrafico, presenza di simboli ed elementi decorativi, e così via). Per essere messe a disposizione degli utenti del web e in particolare di Europeaana senza incorrere in problemi legali, tali immagini devono avere una licenza o di tipo CC0, cioè essere libere da copyright e utilizzabili anche a fini commerciali, o di tipo CCBYSA, essere, cioè, utilizzabili a fini non commerciali e a patto che si citi la fonte da cui sono tratte. Ottenere tali licenze sia dai proprietari delle immagini, che dai proprietari degli oggetti (che a volte coincidono, ma non sempre) non è impresa facile, dal momento che non esiste una normativa europea unica in materia di copyright, anche se sono in corso varie iniziative da parte di Europeaana Foundation per sensibilizzare il Parlamento Europeo su questo tema (cf., ad esempio, <http://pro.europeana.eu/blogpost/eu-parliament-in-favour-of-copyright-rules-better-fit-for-a-digital-age>). Ogni paese si regola diversamente, in modo più o meno restrittivo, e anche nella legislazione di uno stesso paese sono presenti vuoti normativi che prestano il fianco a più di un'interpretazione. Quel che è certo è che anche le norme più avanzate e liberali risultano comunque obsolete di fronte al dilagare dell'uso delle foto digitali determinato dal diffondersi di smartphone e tablet dotati di fotocamera e collegati via web ai diversi sistemi di condi-

visione e diffusione delle immagini (Picasa, Flickr, Wikimedia Commons, senza parlare dei social network come Facebook, Twitter, Instagram, ecc.). Un fenomeno che sarebbe pura utopia cercare di arginare, e che alcuni musei hanno deciso di controllare 'cavalcandolo', cioè mettendo essi stessi online le immagini digitali delle loro collezioni. È il caso del Rijksmuseum di Amsterdam o del Metropolitan di New York, solo per citare un paio degli esempi più famosi, ma le cronache quotidiane registrano sempre nuove adesioni a questa pratica virtuosa (come dimostra la recente iniziativa del Statens Museum for Kunst di Copenaghen che ha deciso di rendere di Public Domain, scaricabili e riutilizzabili gratuitamente, le immagini digitali delle sue collezioni: <http://pro.europeana.eu/blogpost/set-art-free-and-the-rest-will-follow>). Anche l'Italia, che tra i paesi europei è quello, insieme alla Grecia, ad avere la legislazione più restrittiva sull'uso delle immagini del proprio patrimonio culturale, ha recentemente fatto un passo avanti in questo senso, con la circolare del Ministro dei Beni e delle Attività Culturali e del Turismo del marzo 2014 che consente il libero uso, a fini personali e di studio, delle immagini scattate nei musei e nei siti archeologici che dal MIBACT dipendono (il testo della circolare è visibile sulla pagina Documenti del sito <http://www.edr-edr.it>: http://www.edr-edr.it/Documenti/MIBACT_Circolare5_2014.pdf), ma molto resta ancora da fare per tutti quei beni - e sono moltissimi - che fanno capo a enti diversi: civici, provinciali, regionali, ecclesiastici, ecc. La questione degli IPR si presenta, dunque, come uno dei punti più 'sensibili' del progetto, nel senso che mentre procediamo nell'acquisire - quasi sempre, devo dire con successo - tutte le autorizzazioni necessarie, stiamo conducendo anche una serie di studi e di azioni, per le quali è stato istituito un apposito Working Group, volti a sensibilizzare il Parlamento Europeo e i governi nazionali sulla necessità di liberalizzare l'accesso, l'uso e la circolazione delle immagini dei beni culturali. EAGLE, come consorzio, ha aderito a un pronunciamento ufficiale in questo senso di Europeana, che, come si può facilmente comprendere, è molto attiva su questo fronte, e ha dato un piccolo ma concreto segnale istituendo un apposito premio, sia nel 2013 che nel 2014, nell'ambito del concorso fotografico Wiki Loves Monuments Italia. Questo *digital contest* che coinvolge molti paesi si basa proprio su immagini condivise e votate via web di monumenti appositamente 'liberati' da parte degli enti che ne detengono i diritti dai vincoli che normalmente li contraddistinguono, e costituisce di per sé un'occasione di valorizzazione di beni non sempre adeguatamente conosciuti.

Ma l'importanza delle immagini nell'ambito di EAGLE risiede anche in un altro motivo, che rappresenta una delle sfide più difficili, ma anche più interessanti che il progetto si propone. Nell'ambito di un progetto europeo di questo tipo, rendere accessibili le iscrizioni antiche significa non solo metterle a disposizione degli studiosi e degli addetti ai lavori, ma anche aprirle alla conoscenza di un pubblico più vasto, fatto di studen-

ti, insegnanti, turisti, e, più in generale, di cittadini curiosi e interessati che - come stiamo constatando - non mancano. Allargare la piattaforma degli utenti di un contenuto costituito da testi scritti in lingue antiche, spesso frammentari e, per di più, con formule e abbreviazioni che non tutti conoscono, sembra un'impresa quasi impossibile, ma anche in questo la tecnologia può venirci in aiuto. Il team del CNR di Pisa che fa capo a Giuseppe Amato e Fabrizio Falchi, infatti, ha messo a punto, alcuni anni fa, un sistema di riconoscimento per immagini che consente di identificare un oggetto attraverso la fotocamera di uno smartphone, senza bisogno di digitare alcuna informazione. Dopo che il software è stato sperimentato con successo sui monumenti della Toscana, nell'ambito del progetto *Visito* (<http://www.visitotuscany.it>), abbiamo pensato di applicarlo anche alle iscrizioni, e la versione di prova che è stata sviluppata su un campione di circa 17.000 immagini ha dato buoni risultati, per cui è possibile accedere alle informazioni relative a un'epigrafe esposta, ad esempio, in un museo, ma anche riprodotta in una pubblicazione, e di cui siano state preventivamente memorizzate una o, meglio, più immagini, semplicemente inquadrandola con un telefonino dotato della relativa applicazione (per una completa illustrazione di questo processo, cf. Amato et al. 2014).

L'uso del riconoscimento per immagini consente, dunque, di aggirare l'indubbio ostacolo rappresentato dalla difficoltà di 'decifrare' un'iscrizione antica, ma resta l'esigenza di farne conoscere il preciso significato anche ai 'non addetti ai lavori', traducendone il testo in una lingua moderna. Questa esigenza ci è stata fatta notare immediatamente, sin dai colloqui preliminari che abbiamo avuto con i funzionari della Commissione Europea prima ancora della presentazione della domanda di finanziamento, per valutarne le potenzialità di successo. Ed è un'esigenza che, nell'ambito di un progetto nato in ambito accademico, avevamo francamente sottovalutato, semplicemente perché non era tra le priorità di un'iniziativa concepita originariamente come strumento per la ricerca scientifica. Consapevoli dell'importanza che questo aspetto avrebbe avuto nella valutazione della domanda, ma anche dell'impossibilità di soddisfare totalmente questa richiesta, ci siamo posti un obiettivo 'minimo', una specie di 'segnale di buona volontà', impegnandoci a rendere accessibili attraverso il progetto EAGLE le versioni in una lingua moderna di una selezione di testi epigrafici per i quali le traduzioni sono già disponibili perché edite in articoli e altre pubblicazioni.

Per raggiungere questo scopo si è rivelata fondamentale la collaborazione con Wikimedia Italia, che fa parte del consorzio EAGLE, in qualità di *subcontractor*, e che ha messo a frutto l'esperienza maturata in anni di collaborazione con le varie iniziative che fanno capo alla grande 'famiglia' di Wikipedia (Wikimedia Commons, Wikidata, Wikibase, ecc.). I 'wikimediani' hanno messo a punto una piattaforma digitale ad hoc, che è stata chiamata EAGLE Mediawiki, e che si presenta come uno strumento di

lavoro semplice da usare, ma anche molto duttile e facilmente espandibile a seconda delle esigenze (cf., in proposito, Liuzzo et al. 2014). Questa piattaforma consente sia l'inserimento ex novo delle traduzioni già pubblicate, ma non ancora disponibili in formato digitale, sia l'acquisizione automatica di tutte le traduzioni già esistenti in edizioni digitali (come, ad esempio, le *Inscriptions of Roman Tripolitania*, tradotte da Joyce Reynolds e recentemente digitalizzate a cura di Gabriel Bodard e Charlotte Roueché - cf. <http://inslib.kcl.ac.uk/irt2009/>) e in alcune delle banche dati che afferiscono a EAGLE (*Hispania Epigraphica Online*, ad esempio, già prevede, per le proprie epigrafi, traduzioni in spagnolo, cui si stanno ora aggiungendo quelle in portoghese, e lo stesso vale per i testi di *PETRAE*, tradotti in francese, e di *The Last Statues of Antiquity*, tradotti in inglese). L'interesse suscitato da questa iniziativa è testimoniato dall'adesione al consorzio del progetto *Attic Inscriptions Online*, coordinato da Stephen Lambert (<https://www.atticinscriptions.com>, su cui cf. Lambert, McCourt 2014), che ha già riversato nella piattaforma EAGLE Mediawiki tutte le sue traduzioni in inglese delle iscrizioni greche dell'Attica. Ma il risultato sta andando ben oltre le nostre aspettative. La possibilità di avere a disposizione per la prima volta all'interno di un unico progetto un numero così elevato di traduzioni di iscrizioni sta portando a una riflessione, forse troppo a lungo rimandata, sull'importanza e la difficoltà del lavoro del traduttore come parte integrante e non irrilevante dell'attività scientifica. Lo testimoniano le relazioni presentate sull'argomento in occasione del primo convegno internazionale EAGLE (Bigi 2014; Campedelli 2014), e il fattivo interesse a collaborare su questo fronte da parte, ad esempio, di studiosi come Leopoldo Gamberale, da anni impegnato, anche come traduttore, nello studio dei *Carmina Latina Epigraphica* (Gamberale 1993, i cui spunti sono stati ripresi e approfonditi nel corso del Workshop *Traduzione a fronte - tradurre iscrizioni tra divulgazione e ricerca*, organizzato a Roma il 25 giugno 2014). Ma la piattaforma EAGLE Mediawiki ha anche delle notevoli potenzialità didattiche nell'ambito dell'insegnamento delle lingue classiche nella scuola secondaria superiore, che nel lavoro su testi non 'canonici' (e quindi non facilmente disponibili in rete già tradotti), nel connubio tra antichità e nuove tecnologie e in un più stretto rapporto tra il mondo della scuola e quello dell'università, può trovare indubbi spunti di interesse (tra i più recenti contributi su questo argomento si vedano, ad esempio, Balbo 2014 e Reali, Turazza 2015). Lo stiamo già sperimentando con Enrico Zuddas, docente di latino e greco al liceo Properzio di Assisi, con Lavinio Del Monaco, del liceo Eliano di Palestrina, e con Paola Tassini, del liceo Tasso di Roma, che hanno coinvolto le loro classi nella traduzione di alcune iscrizioni della loro città, ma contiamo di farlo anche con altri insegnanti che hanno un particolare interesse e una specifica competenza in ambito epigrafico. Naturalmente, anche per questa parte del progetto, non viene meno l'impegno di EAGLE a mettere online mate-

riali di qualità controllata, per cui di ogni traduzione, se già edita, viene dichiarato autore, luogo e data di pubblicazione, mentre se è opera di uno studente, viene indicato l'insegnante che si assume la responsabilità della sua supervisione. Per tutte, poi, è stato concepito un sistema di revisione e controllo che coinvolge il progetto Perseids, con sede presso la Tufts University, attualmente impegnato nell'elaborazione di un modello di edizione digitale partecipata, applicabile anche ai testi epigrafici, che rende la sua collaborazione con EAGLE particolarmente interessante.

Come si può vedere, di strada ne è stata fatta molta da quel lontano 2003 che ha visto la nascita della federazione EAGLE, e molta ne resta ancora da fare per poter dire di aver realizzato i propositi che ci proponiamo. Non mi resta che augurarmi che chi presenterà il progetto in occasione della prossima edizione di questo seminario possa illustrare non solo obiettivi, ma anche risultati, nella consapevolezza, comunque, che ogni traguardo, in questo come in altri campi, sarà sempre un nuovo punto di partenza.

Bibliografia

- Amato, Giuseppe; Falchi, Fabrizio; Rabitti, Fausto; Vadicamo, Lucia (2014). «Inscriptions Visual Recognition. A Comparison of State-of-the-Art Object Recognition Approaches». Casarosa et al. 2014, 117-31.
- Balbo, Andrea (2014). «Alcune possibilità per la didattica multimediale del latino. I siti web, le videopresentazioni, le eventuali risorse di YouTube». *Nuova Secondaria Ricerca*, 5, 44-51.
- Bigi, Francesca (2014). «Towards an EAGLE Standard in Translating Inscriptions». Casarosa et al. 2014, 179-86.
- Campebelli, Camilla (2014). «Translating Greek and Roman Inscriptions». Casarosa et al. 2014, 179-86.
- Casarosa, Vittore; Liuzzo, Pietro Maria; Orlandi, Silvia; Santucci, Raffaella (eds.) (2014). *Information Technologies for Epigraphy and Cultural Heritage = Proceedings of the First EAGLE International Conference*. Roma: Sapienza Università Editrice. Collana Convegni 26.
- Casarosa, Vittore et al. (2014). «A Conceptual Model for Inscriptions. Harmonizing Digital Epigraphy Data Sources». Casarosa et al. 2014, 23-40.
- Eck, Werner; Funke, Peter; Dohnicht, Marcus et al. (2014). *Öffentlichkeit – Monument – Text. XIV Congressus Internationalis Epigraphiae Graecae et Latinae 27.-31. Augusti MMXII = Akten*. Berlin: De Gruyter.
- Felle, Antonio Enrico (2012). «Esperienze diverse e complementari nel trattamento digitale delle fonti epigrafiche. Il caso di EAGLE ed EpiDoc». *Diritto romano e scienze antichistiche nell'era digitale = Convegno di studio* (Firenze, 12-13 settembre 2011). Torino: Giappichelli, 117-30.
- Felle, Antonio Enrico (2014). «Prospettive per il corpus digitale delle iscrizioni cristiane di Roma: contesti e testi». Eck et al. 2014, 508-10.

- Gamberale, Leopoldo (1993). «Letteratura minima. I Carmina Latina Epigraphica». *Cultura e lingue classiche 3 = II Convegno di aggiornamento e di didattica* (29 ottobre-1 novembre 1989, Palermo). A cura di Biagio Amata. Roma: L'Erma Di Bretschneider, 379-403.
- Giberti, Luca Marco Carlo; Orlandi, Silvia; Santucci, Raffaella (2014). «EAGLE – Europeana Network of Ancient Greek and Latin Epigraphy. Making the Ancient Inscriptions Accessible». *Lexicon Philosophicum*, 2. URL <http://lexicon.cnr.it/index.php/LP/article/view/408> (2017-10-17).
- Hainzmann, Manfred (2014). «Monumenta Epigraphica Electronica – Plädoyer für eine transmediale Edition lateinischer Inschriften». Eck et al. 2014, 510-12.
- Lambert, Stephen; McCourt, Frank (2014). «Attic Inscriptions Online (AIO). Attic Inscriptions in English Translation». Casarosa et al. 2014, 155-63.
- Liuzzo et al. (2014). «The EAGLE Mediawiki. A Fully Collaborative Database for Academics, Data Engineers and the General Public». Casarosa et al. 2014, 187-99.
- Pancierà, Silvio (2006). «What Do We Need?». *Epigrafi, epigrafia, epigrafisti. Scritti vari, editi e inediti (1956-2005) con note complementari e indici*. Roma: Edizioni Quasar, 1917-8.
- Pancierà, Silvio (2012). «Epigraphy and Informatics. An Introduction». *Epigraphy and Historical Sciences*. Oxford: Oxford University Press, 271-3.
- Pancierà, Silvio (2013). «Notizie da EAGLE». *Epigraphica*, 75, 502-6.
- Reali, Mauro; Turazza, Gisella (2015). «Parole di pietra. Epigrafia e didattica del latino». Balbo, Andrea; Ricucci, Marco (eds.), *Prospettive per l'insegnamento del latino*. Torino: Loescher Editore, 45-57. I Quaderni della Ricerca 16.

Alcune note di ecdotica digitale nella prospettiva della filologia collaborativa e cooperativa

Federico Boschetti

(Istituto di Linguistica Computazionale "A. Zampolli", CNR-ILC, Pisa, Italia)

Abstract In the digital age, classical philology is becoming more and more collaborative and cooperative. Even if these terms are apparently synonyms, the distinction adopted by pedagogues in the last decades is useful to describe different ways to organise collective works. Collaboration aims at the realisation of a final product, which is the result of negotiations among the involved parts. Cooperation aims at the creation of independent subproducts, which will be coordinated to achieve the agreed final result, but which can also be assembled in unpredictable new ways by third parties. In order to integrate these approaches in the domain of classical philology, first of all it is necessary to clarify the needs of a typical philologist, not just the requirements of the user of a specific project. For this reason, the last part of the article is devoted to a case study that takes into account the exploration of the secondary literature, often neglected by digital philologists but crucial to the ecdotic best practices.

Sommario 1 Introduzione. – 2 Integrazione fra filologia collaborativa e filologia cooperativa. – 3 La specializzazione nell'ambito della filologia cooperativa. – 4 La peculiarità della filologia cooperativa rispetto alla linguistica computazionale. – 5 Dall'interscambio all'interoperabilità. – 6 Il coordinamento dei mezzi minimi per l'esercizio della filologia cooperativa. – 7 Discussione di un caso d'uso. – 8 Conclusione.

Keywords Classical philology. Collaborative and cooperative philology. Digital philology. Ecdotics. Perseus Digital Library.

1 Introduzione

L'edizione scientifica digitale si sta progressivamente differenziando dall'edizione a stampa, in quanto si configura sempre più come un insieme di viste su un processo aperto anziché come un singolo prodotto in sé concluso. Nell'era della cosiddetta quarta rivoluzione (Roncaglia 2010), il trasferimento dal supporto cartaceo ai supporti digitali aumenta inevitabilmente il rischio di una deriva testuale e per questo la riflessione teorica sulle pratiche editoriali deve accompagnare sia l'opera di digitalizzazione delle risorse già esistenti, sia la creazione di risorse digitali native (Pierazzo 2014).

2 Integrazione fra filologia collaborativa e filologia cooperativa

Negli studi filologici, l'evoluzione della rete telematica in social network ha comportato l'apertura della scrivania del singolo studioso alla comunità dei pari non più soltanto per fruire del prodotto finale, licenziato dall'autore o dall'editore, ma anche per intervenire in corso d'opera. La filologia nell'era digitale è chiamata quindi a essere collaborativa e cooperativa (per la distinzione fra i due termini, si veda Kozar 2010). La collaborazione prevede che una comunità con obiettivi condivisi coordini i propri sforzi per la realizzazione di un prodotto, come ad esempio l'edizione critica di un'opera antica. La cooperazione invece permette ai membri o ai gruppi di diverse comunità di perseguire i propri obiettivi specifici, usando i servizi offerti dagli altri e offrendo in cambio i propri, come ad esempio il servizio di analisi metrica e il servizio di lemmatizzazione che, combinati da terze parti, possono produrre l'analisi formulare, anche se non pianificata inizialmente dai fornitori dei servizi originari.

La filologia cooperativa promuove quindi il disaccoppiamento delle funzionalità offerte dagli obiettivi specifici che devono essere raggiunti. Questo comporta il passaggio dallo sviluppo di applicazioni monolitiche, nate per un singolo progetto, alla creazione da un lato di archivi digitali codificati in modo standard e annotabili da terze parti e dall'altro allo sviluppo di librerie di componenti e di servizi web, che possono essere riusati per nuovi progetti di ricerca (Boschetti, Del Grosso 2015). L'approccio collaborativo e l'approccio cooperativo non sono mutualmente esclusivi, ma anzi si integrano reciprocamente. Infatti, le risorse digitali o gli strumenti computazionali realizzati in seno ad un progetto collaborativo amplificano la loro utilità quando sono impiegati all'interno di altri progetti realizzati da terze parti.

Ad esempio, sul versante delle risorse, la cooperazione fra i progetti collaborativi *Musisque Deoque* e *Perseus Project* permette agli studiosi di fruire dei testi latini rinascimentali sia attraverso l'interfaccia originaria che attraverso l'interfaccia della *Perseus Digital Library*. L'interfaccia originaria (<http://www.mqd.it>) è nata per lo studio dell'intertestualità e mette a disposizione un motore di ricerca avanzato e un sistema di visualizzazione che facilita la comparazione dei testi. D'altro canto, l'interfaccia della *Perseus Digital Library* (<http://www.perseus.tufts.edu>) mostra la sua massima efficacia per lo studio linguistico, grazie all'associazione di ciascuna parola alla relativa analisi morfologica e alla relativa voce lessicale proveniente da dizionari bilingui consultabili online. A loro volta, i testi della *Perseus Digital Library* sono accessibili anche attraverso *Perseus Under PhiloLogic*, grazie all'interfaccia sviluppata presso l'Università di Chicago (<http://perseus.uchicago.edu>), che potenzia ulteriormente i servizi di visualizzazione delle analisi linguistiche.

Un secondo esempio, questa volta sul versante degli strumenti e dei servizi, è l'accesso tramite *web service* alle analisi prodotte dallo strumento di

lemmatizzazione latina sviluppato presso l'Istituto di Linguistica Computazionale "A. Zampolli" (end point: <http://cophilab.ilc.cnr.it:8000/aucepsws/morphanalysis/j/> seguito dalla forma flessa latina da analizzare, ad es. cano). Nel fornire questo servizio, si è seguita la strada dell'*Alpheios Project*, che ha creato dei web services per la lemmatizzazione delle lingue storiche (<https://github.com/PerseusDL/morpheus-wrapper>) a partire dall'originario sistema di analisi morfologica *Morpheus* (Crane 1991).

3 La specializzazione nell'ambito della filologia cooperativa

Nella fase pionieristica e artigianale dello sviluppo di software per le *Digital Humanities*, il filologo per lo più interagiva con un singolo programmatore oppure creava autonomamente i propri programmi, facendo convergere tutte le competenze necessarie in una o al massimo due figure professionali. Nella fase matura delle *Digital Humanities*, dove più individui e più gruppi di diverse comunità interagiscono per la creazione di prodotti digitali, la specializzazione si rende necessaria. In particolare, è bene tener distinte le competenze richieste per l'elaborazione di modelli formali (ad esempio, per la modellazione della struttura dati più efficaci a rappresentare la variantistica), per la creazione di risorse digitali (ad esempio, l'archivio di *Musisque Deoque*), per lo sviluppo di strumenti computazionali (ad esempio, gli algoritmi di *Tesserae* per lo studio dell'intertestualità: <http://tesserae.caset.buffalo.edu>) e per la costruzione di infrastrutture (ad esempio, la *Perseus Digital Library*). La filologia cooperativa è costituita quindi dai seguenti sottodomini: formale, digitale, computazionale e infrastrutturale (o *e-philology*). I modelli formali degli oggetti e dei processi coinvolti nella disciplina filologica prescindono dalle contingenze delle tecnologie disponibili a implementarli e dirigono le modalità di codifica digitale, di analisi computazionale e di gestione infrastrutturale.

Un esempio dovrebbe chiarire e anticipare quanto sarà esposto nelle sezioni seguenti. I manoscritti da un lato e le collazioni dall'altro sono rispettivamente oggetti e processi dell'ambito filologico che devono essere formalizzati a prescindere dal livello tecnologico raggiunto per acquisire immagini ad alta definizione o dallo specifico linguaggio di programmazione con cui è implementato un algoritmo per la collazione automatica. A livello formale, un manoscritto deve essere descritto tramite le relazioni delle sue parti costituenti e la collazione deve essere descritta tramite la sequenza delle operazioni elementari di cui è composta. La codifica digitale del manoscritto, la implementazione dell'algoritmo per la collazione automatica di più manoscritti e la creazione dell'infrastruttura web che permette l'interazione fra oggetti e processi si devono conformare a tali modelli astratti.

4 La peculiarità della filologia cooperativa rispetto alla linguistica computazionale

Fin dalle origini con padre Busa, la ricerca filologica in ambito digitale ha mutuato strumenti e metodi dalla linguistica computazionale. Se da un lato questo ha permesso di analizzare testi di interesse storico e letterario sfruttando i risultati delle ricerche in campo linguistico, dall'altro ha frenato lo sviluppo di strumenti e metodi propri dell'ambito filologico.

La parte della linguistica che più si avvicina agli interessi del filologo è la *Corpus Linguistics* (McEnery, Wilson 1996), in quanto quest'ultima concentra la sua attenzione sui fenomeni della lingua che si possono estrarre da collezioni di testi reali e non da insiemi di esempi fittizi. Alla sua natura descrittiva, la *Corpus Linguistics* aggiunge inoltre la componente quantitativa, per cui i fenomeni non vengono soltanto individuati, ma anche conteggiati e pesati, garantendo un alto livello di oggettività e di falsificabilità ai risultati raggiunti.

Tuttavia il linguista non è interessato alla materialità e ai contesti originari delle fonti primarie da cui i suoi testi derivano, non è interessato cioè agli originari dispositivi di comunicazione (per il concetto, si veda Lamé 2015) che hanno veicolato i testi. Inoltre per il linguista il testo è dato, non è esso stesso oggetto di discussione ed è quindi un punto di partenza, non di arrivo. Le collezioni di testi che interessano il linguista provengono da edizioni singole, scelte con criteri assiologici. Infine, i molteplici livelli di analisi (morfosintattica, semantica, tematica, ecc.) applicati al testo prevedono una sola interpretazione e, nel caso in cui più annotatori o più sistemi automatici di analisi linguistica forniscano dati divergenti, il risultato finale viene armonizzato da un supervisore umano o da un sistema di valutazione dell'analisi più probabile.

Per il filologo, al contrario, questi tre punti di articolazione sono sostanziali per la sua disciplina: molteplicità di dispositivi di comunicazione che veicolano una molteplicità di varianti testuali, su cui insiste una molteplicità di interpretazioni distribuite su diversi livelli di analisi. Di conseguenza, le molteplicità che interessano al filologo non possono essere affrontate solo a livello di codifica dei testi, ma vanno affrontate anche a livello di strumenti di analisi e di infrastrutture telematiche.

5 Dall'interscambio all'interoperabilità

Le risorse digitali, e in particolare i corpora testuali, sono la materia prima della filologia cooperativa. È comprensibile quindi che l'attenzione degli umanisti digitali si sia concentrata soprattutto sull'acquisizione e la codifica delle risorse, dove si sono raggiunti da un lato alti livelli di accuratezza del testo tramite avanzati sistemi di OCR, e dall'altro elevati livelli

di standardizzazione della codifica grazie alla *Text Encoding Initiative*.

Tuttavia, la modellazione dei dati e la standardizzazione dei formati, pur necessaria all'interscambio delle risorse, non garantisce l'interoperabilità fra i processi che gestiscono i dati (Bauman 2011). Se si pensa ad esempio alla codifica di un apparato critico secondo le molteplici opzioni offerte da TEI (<http://www.tei-c.org>), si può notare quanto sia facile, grazie ad appositi fogli di stile, visualizzare le notizie d'apparato in modo simile a un'edizione critica a stampa, ma quanto sia invece difficile eseguire operazioni più complesse, come l'allineamento delle singole parole di una variante sulle parole del testo accolto dall'editore. Attualmente, al di là delle trasformazioni eseguibili tramite fogli di stile da un documento a un altro documento, la gestione dei dati codificati in TEI è per lo più affidata agli sviluppatori di applicazioni specifiche.

Per questo motivo la comunità delle *Digital Humanities* si sta orientando verso forme più consapevoli di interoperabilità, e il progetto europeo *Interedition* è uno dei testimoni più rappresentativi di questa tendenza (Haentjens Dekker et al. 2015, 452):

Interoperability is the key term within the framework of [...] Interedition, whose aim is 'to encourage the creators of tools for textual scholarship to make their functionality available to others, and to promote communication between scholars so that we can raise awareness of innovative working methods'. The tools developed by Interedition's 'Prototyping' working group were tested by other research teams, which formulate strategic recommendations.

6 Il coordinamento dei mezzi minimi per l'esercizio della filologia cooperativa

Come si è visto, la filologia cooperativa si proietta su uno scenario dove le crescenti risorse digitali sono interoperabili e dove gli strumenti computazionali complessi sono costruiti tramite l'assemblaggio di componenti semplici e la concatenazione di servizi web.

A livello delle risorse, come più volte rimarcato da Crane (si veda, ad esempio, Crane et al. 2009), il punto di partenza per la cooperazione è la condivisione di uno schema citazionale (messo a disposizione dai *Canonical Text Services*, CTS) in grado di identificare in modo inequivocabile le unità testuali della fonte primaria a qualsiasi livello di granularità (libro, capitolo, parola, singolo carattere). In questo modo, le molteplici annotazioni (non solo metriche, morfosintattiche, semantiche, ma anche storiche od estetiche, ad esempio) che insistono su una particolare versione del testo, sono totalmente disaccoppiate dal documento che rappresenta la fonte primaria e fanno ad esso riferimento solo attraverso lo schema citazionale,

che rende ragione anche della specifica versione del documento. Dato che ciò che viene citato è indipendente da ciò che lo cita, la quantità di informazioni associate e portate da terze parti può crescere in modo indefinito.

A livello degli strumenti, i servizi di analisi linguistica tramite *web services* stanno rivelando la loro efficacia, mentre ancora molto deve essere fatto per la creazione di librerie di componenti software con cui costruire i *web services* stessi, le *web*, *desktop*, e *mobile applications* senza dover reimplementare tutte le funzionalità con tecnologie sempre diverse per le diverse destinazioni.

Il paradosso che rallenta maggiormente lo sviluppo della filologia cooperativa è la scarsa comunicazione fra la comunità delle *Digital Humanities* e la comunità dei filologi tradizionali. I contatti non sistematici fra le due comunità sono la causa di analisi dei requisiti e specifiche parziali, legate a visioni basate sulle esigenze di singoli progetti anziché sulle ampie possibilità offerte dal dominio della filologia nel suo complesso.

7 Discussione di un caso d'uso

Non vi è ancora sufficiente interoperabilità fra gli strumenti creati per lo studio delle fonti primarie dirette (ad esempio, tramite collazioni automatiche), lo studio delle fonti indirette (ad esempio, tramite estrazione di citazioni e allusioni) e, infine, lo studio della storia della tradizione (ad esempio, tramite lo studio della fortuna di varianti antiche).

Il caso d'uso scelto prende in considerazione un ulteriore aspetto di cui l'editore deve tener conto, che sta emergendo anche grazie alla grande disponibilità di fonti secondarie digitalizzate (commentari, articoli, monografie, ecc.): la storia degli studi filologici.

Per esempio, al v. 135 dei *Persiani* di Eschilo la variante ἀκροπενθεῖς è attestata da tutti i codici tranne Y, che reca ἄβροπενθεῖς. È ovvio che, per una tradizione complessa come quella eschilea, l'apparato digitale non si può limitare alle sole informazioni derivanti dalla collazione di manoscritti. La variante ἄβροπενθεῖς è accolta dalla maggior parte degli editori dopo Paley non per il valore di Y, ma per le ragioni di ordine idiomatico, stilistico, paleografico e di evidenza di uno *scholium* (che contiene il verbo ἀβρύνεσθαι) addotte dallo stesso Paley, che non conosceva Y e che quindi propose ἄβροπενθεῖς per congettura. Di più: lo studio approfondito del *modus operandi* di Paley (Richardson 1957) ci informa che la sua congettura non nasce dall'evidenza estrinseca dello *scholium*, ma che in esso si rafforza semmai l'ipotesi già formulata grazie alle altre ragioni, intrinseche.

Un'edizione scientifica digitale deve coordinare i mezzi per codificare anche questo tipo di informazioni provenienti dalla letteratura secondaria (il mesotesto di J. Cummings, menzionato in Pierazzo 2015) e i mezzi per reperire e visualizzare queste informazioni.

8 Conclusione

In questo breve contributo si è cercato di evidenziare quali siano i principi della filologia cooperativa e come il concetto di interoperabilità non si possa limitare alla codifica delle risorse digitali, ma debba prendere in considerazione gli strumenti computazionali e le infrastrutture telematiche che insistono su tali risorse.

Bibliografia

- Bauman, Syd (2011). «Interchange vs. Interoperability». *Proceedings of Balisage = The Markup Conference 2011* (Montréal, 2nd-5th August 2011). Balisage Series on Markup Technologies 7. URL <https://www.balisage.net/Proceedings/vol7/cover.html> (2017-11-27).
- Boschetti, Federico; Del Grosso, Angelo Mario (2015). «TeiCoPhiLib. A Library of Components for the Domain of Collaborative Philology» [online]. *Journal of the Text Encoding Initiative*, 8. URL <https://jtei.revues.org/1285> (2017-10-17).
- Crane, Gregory (1991). «Generating and Parsing Classical Greek». *Literary and Linguistic Computing*, 6(4), 243-5.
- Crane, Gregory; Seales, Brent; Terras, Melissa (2009). «Cyberinfrastructure for Classical Philology» [online]. *Digital Humanities Quarterly*, 3(1). URL <http://www.digitalhumanities.org/dhq/vol/3/1/000023/000023.html> (2017-10-17).
- Haentjens Dekker, Ronald et al. (2015). «Computer-Supported Collation of Modern Manuscripts. Collatex and the Beckett Digital Manuscript Project». *Digital Scholarship in the Humanities*, 30(3), 452-70.
- Kozar, Olga (2010). «Towards Better Group Work. Seeing the Difference between Cooperation and Collaboration». *English Teaching Forum*, 48(2), 16-23.
- Lamé, Marion (2015). «Primary Sources of Information, Digitization Processes and Dispositive Analysis». Tomasi, Francesca; Rosselli Del Turco, Roberto; Tammara, Anna Maria (a cura di), *Proceedings of the Third AIUCD Annual = Conference on Humanities and their Methods in the Digital Ecosystem*. New York: ACM, 1-5.
- McEnery, Tony; Wilson, Andrew (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Pierazzo, Elena (2015). *Digital Scholarly Editing. Theories, Models and Methods*. Farnham (Surrey): Ashgate.
- Roncaglia, Gino (2010). *La quarta rivoluzione. Sei lezioni sul futuro del libro*. Roma-Bari: Laterza.
- Richardson, L.J.D. (1957). «Aeschylus Persae 135. ἀβροπενθεϊς Paley (codd. ἀκροπενθεϊς)». *Hermathena*, 89, 73-5.

Le sfide di una biblioteca digitale del latino tardoantico

Struttura, canone e questioni aperte di codifica e visualizzazione

Alice Borgna

(Università del Piemonte Orientale, Italia)

Simona Musso

(Università del Piemonte Orientale, Italia)

Abstract This paper focuses on different aspects of the *DigilibLT* project (Digital Library of Late Antique Latin Text), illustrating not only its functionality, but also problems and difficulties in terms of project building and management. To create a digital library, the indisputable starting point is the definition of the canon of the texts to be included; in this sense, our selection, based on late antique Latin texts, pagan and in prose, revealed some critical point. First of all, since most of these works have technical-scientific or educational purposes, they are in many cases accompanied by diagrams, images and symbols, a graphic apparatus that often can be made digital only with some difficulties. Besides, on a more general level, we will also explain how the strong participation of young people in the team has played a crucial role in the development of the project; it has also contributed to enrich the CV of plenty humanist graduates with significant competencies in the field of the Digital Humanities.

Sommario 1 Costruire una biblioteca digitale. – 1.1 Consultazione e *download*. – 1.2 La ricerca. – 1.3 Un ambiente di lavoro: schede e bibliografia. – 2 Canone. – 3 Formati e visualizzazione. – 4 Glifi. – 5 Natura collaborativa e qualità: un rapporto problematico. – 6 Digitale, perché? – 6.1 Un punto di partenza.

Keywords Digital Libraries. Project building. XML-TEI encoding. Glyphs. Late antiquity.

La rivoluzione informatica che ha investito tutti i campi di studio, anche quelli che potevano apparire maggiormente legati a strumenti tradizionali, come la filologia classica, negli ultimi decenni sta conoscendo un'evoluzione altrettanto importante. Molte delle banche dati testuali nate su CD-ROM, infatti, oggi più o meno ovunque vengono riversate sul web, spesso

con qualche difficoltà,¹ a dimostrazione di un problema, quello della rapida obsolescenza del supporto tecnologico, che rappresenta una delle grandi sfide della contemporaneità. Non solo: la dimensione online di database e biblioteche digitali suscita anche questioni sostanziali, legate all'accesso libero e pone in primo piano la necessità che risorse affini trovino il modo di coordinarsi e collaborare.

Questi rapidi sviluppi hanno anche reso urgente la sistemazione digitale di quel materiale che ne è ancora privo, una lacuna che può condizionare fortemente anche lo stesso progredire della ricerca scientifica. Era questo il caso della letteratura latina tardoantica che, fino a qualche anno fa, la rete offriva in forme sporadiche e non organizzate, ovvio riflesso di una minor fortuna, anche a livello cartaceo, dell'approfondimento critico in merito alle opere dei secoli II-VI d.C.² Con l'intento di ovviare a questa assenza, nel 2010 è nato il progetto *Digital Library of Late Antique Latin Texts* (digilibLT), che ha costruito e via via implementato una biblioteca digitale dedicata in modo particolare alla prosa pagana.³

1 Costruire una biblioteca digitale

1.1 Consultazione e download

Premessa necessaria alla costruzione di una risorsa digitale è stata, naturalmente, una prima stesura di un canone di riferimento che permettesse l'individuazione di tutte le opere e di tutti gli autori da comprendere.⁴ Questo canone, di circa quattrocento voci, non è solo scaricabile come

1 È questo ad esempio il caso, su cui vi sarà modo di tornare, del prototipo di CD-ROM (*GL-CD*) contenente il corpus delle opere grammaticali. Cf. *infra* § 3.

2 Lo stesso *Thesaurus linguae Latinae*, con il quale *DigilibLT* ha avuto modo di comunicare (cf. *infra* § 6), non offre una schedatura completa delle opere dalla metà del II secolo in poi, soprattutto quelle di contenuto profano. Per lo stato frammentario e disomogeneo in cui la rete offriva i testi della letteratura latina tardoantica prima di *DigilibLT*, cf. Tabacco 2014, 3-16, Borgna 2017, 4-5.

3 URL www.digiliblt.uniupo.it. Il progetto è stato finanziato dalla Regione Piemonte in seguito a un bando competitivo finalizzato a ricerche nell'ambito delle Scienze umane e sociali. Il gruppo di ricerca, che fa capo all'Università del Piemonte Orientale, è diretto da Raffaella Tabacco, responsabile della ricerca e condirettore per la parte relativa alle competenze disciplinari di Letteratura latina, Maurizio Lana, a capo delle competenze di *Digital Humanities*. L'Università di Torino è stata proponente aggiuntivo e ha visto l'installazione di una seconda unità di ricerca presso il Dipartimento di Studi Umanistici con Ermanno Malaspina e Andrea Balbo. Per i molteplici aspetti del progetto, anche informatici, si vedano Lana 2011, 2012b, 2012c; Tabacco 2014, Mosca-Musso 2016; Borgna 2017.

4 Cf. *infra* § 3.

file autonomo, ma è anche attivamente ricercabile nel sito per autore, per opera, per ordine alfabetico o per secolo. Si riporta, come esempio, la schermata che si ottiene ricercando gli autori che iniziano con la lettera P: nell'elenco compaiono, insieme con gli autori, anche le opere anonime (in corsivo).

digilibLT Biblioteca digitale di testi latini tardoantichi

UPO Fondazione Casa di Riposo di Veretti REGIONE PIEMONTE

home | Il progetto | Notizie | Tardoantico nel web | Aiuto | Contatti e feedback | Accedi |

Ricerca

CERCA

Opere disponibili per la ricerca
Ricerca avanzata
Ricerca bibliografica

Letture e download di testi

Opere
 Autori
 Bibliografia

Per ordine cronologico
II III IV V VI VII VIII
Datazione incerta
Elenco opere

Per ordine alfabetico
A B C D E F G H I L M N O P
Q R S T V
Elenco opere

Risorse
Opere sul tardoantico
Canone autori tardoantichi
Download di font e software
[Scarica tutte le opere](#)

Proponente aggiuntivo

Scegliere nella parte di sinistra l'ordinamento:
Cronologico (per secolo, datazione incerta o elenco completo)
Alfabetico (per lettera iniziale o elenco completo)

Opere che iniziano con la lettera P

1. Panegyrici Latini - saec. III-IV
2. Panegyricus dictus Augustoduni a. 297-8 - oratio pro instaurandis scholis - saec. III
3. Panegyricus dictus Constantino filio Constantii (Treueris a. 313 p. Chr.) - saec. IV
4. Panegyricus dictus Constantino imperatori (Romae a. 321 p. Chr.) - saec. IV
5. Panegyricus dictus Constantino imperatori (Treueris a. 310 p. Chr.) - saec. IV
6. Panegyricus dictus Constantino imperatori (Treueris a. 312 p. Chr.) - Gratiarum actio - saec. IV
7. Panegyricus dictus Constantio Caesari (Treueris a. 297-8) - saec. III
8. Panegyricus dictus Iuliano imperatori (Constantinopoli a. 362 p. Chr.) - Gratiarum actio de consulatu suo - IV saec.
9. Panegyricus dictus Maximiano et Constantino (Treueris ? a. 307 p. Chr.) - saec. IV
10. Panegyricus dictus Maximiano et Diocletiano (Treueris a. 289 p. Chr.) - saec. III
11. Panegyricus dictus Theodosio imperatori (Romae a. 389 p. Chr.) - saec. IV
12. Panegyricus dictus Treueris ? a. 291 p. Chr. - Genethliacus Maximiani Augusti - saec. III
13. *Partitiones - saec. V-VI*
14. Passio Sanctarum Perpetuae et Felicitatis - 203 ca.
15. *Pauli Diaconi Epitome - saec. II (VIII)*
16. Peregrinatio Egeriae - 384 ca.
17. Peri hermeneias liber Aristotelis latine uersus - saec. VI
18. Periochae Liuii - saec. III-IV
19. Philosophiae consolatio - 523-524
20. *Physiologus latinus rec. B - saec. IV-V*

Figura 1. Autori e opere anonimi che iniziano con la lettera P

Ancora più ricca di informazioni è la ricerca per opera (fig. 2), che segna con apposite icone la presenza del testo all'interno della banca dati, precisando in quale formato essa sia disponibile. I testi, infatti, oltre ad consultabili online, sono anche scaricabili gratuitamente (previa una registrazione a soli fini statistici) e per intero in vari formati (.txt, TEI, PDF,

ePub), in modo che ogni studioso possa utilizzarli nel modo più adeguato ai suoi interessi di ricerca o di lettura.⁵ Tutte le opere sono poi distribuite su licenza Creative Commons (cf. Lana 2012a, 48-64), nel rispetto della filosofia *open source* e in modo da rendere il lavoro già fatto la base di partenza per altra ricerca (cf. *infra* § 7.2).

The screenshot shows the digilibLT website interface. At the top, there is a navigation bar with the logo 'digilibLT' and the text 'Biblioteca digitale di testi latini tardoantichi'. To the right of the logo are logos for 'UPO' (Università del Piemonte Orientale), 'Fondazione Cassa di Risparmio di Novara', and 'REGIONE PIEMONTE'. Below the navigation bar, there is a search bar with a 'CERCA' button. On the left side, there are several filters: 'Opere disponibili per la ricerca', 'Ricerca avanzata', 'Ricerca bibliografica', 'Lettura e download di testi' (with radio buttons for 'Opere', 'Autori', and 'Bibliografia'), 'Per ordine cronologico' (with a list of centuries II, III, IV, V, VI, VII, VIII), 'Datazione incerta', 'Elenco autori', 'Per ordine alfabetico' (with a list of letters A, B, C, D, E, F, G, H, I, L, M, N, O, P, Q, R, S, T, V), and 'Elenco autori'. Below these filters, there is a 'Risorse' section with links for 'Opere sul tardoantico', 'Canone autori tardoantichi', 'Download di font e software', and 'Scarica tutte le opere'. At the bottom left, there is a 'Proponente aggiuntivo' section with a circular logo. The main content area displays a list of authors and works starting with the letter 'P', titled 'Autori e opere anonime che iniziano con la lettera P'. The list includes: 1. Palaemon (Ps.), 2. Palaemon (ps.) = Victorinus (Ps.) - saec. VI / VII, 3. Palladius - saec. V, 4. Panegyricus dictus Constantino filio Constantii (Treueris a. 313 p. Chr.) - saec. IV, 5. Panegyricus dictus Constantino Imperatori (Treueris a. 310 p. Chr.) - saec. IV, 6. Panegyricus dictus Constantino imperatori (Treueris a. 312 p. Chr.) - Gratiarum actio - saec. IV, 7. Panegyricus dictus Constantio Caesari (Treueris a. 297-8) - saec. III, 8. Panegyricus dictus Maximiano et Constantino (Treueris ? a. 307 p. Chr.) - saec. IV, 9. Papiir(an)us, 10. Passio Sanctarum Perpetuae et Felicitatis - 203 ca., 11. Pelagonius - saec. IV (sec.), 12. Peregrinatio Egeriae - 384 ca., 13. Periochae Liuii - saec. III-IV, 14. Philargyrius (Ps.) - saec. VII, 15. Philargyrius/Philagrius - saec. V, 16. Phocas - saec. V in., 17. Phocas (Ps.), 18. Physiologus latinus rec. B - saec. IV-V, 19. Physiologus latinus rec. Y, 20. Placitus - saec. V (dub.). At the bottom of the list, there is a page number '1 2 »'.

Figura 2. Situazione al 2015-06-25

Un elenco preciso di tutte le opere disponibili per la ricerca e il download (divise per secoli) al momento in cui si consulta la biblioteca digitale è fornito dal primo pulsante sulla colonna di sinistra. Ogni nuova accessione è poi segnalata in home page da un'apposita notizia. La figura 3 riporta, come esempio, l'elenco delle opere del V secolo a disposizione dell'utente.

5 Sulle ragioni per cui si è deciso di abbandonare il formato Beta Code, ormai destinato all'obsolescenza, cf. Tabacco 2014, 10 n. 21.

digilibLT Biblioteca digitale di testi latini tardoantichi

UPO FONDAZIONE UNIVERSITÀ DEL PIEMONTE ORIENTALE

FONDAZIONE REGIONE PIEMONTE

home | Il progetto | Notizie | Tardoantico nel web | Aiuto | Contatti e feedback | Accedi

Ricerca

Opere disponibili per la ricerca

Ricerca avanzata

Ricerca bibliografica

Letture e download di testi

Opere

Autori

Bibliografia

Per ordine cronologico

II III IV V VI VII VIII

Datazione incerta

Elenco opere

Per ordine alfabetico

A B C D E F G H I L M N O P

Q R S T V

Elenco opere

Risorse

Opere sul tardoantico

Canone autori tardoantichi

Download di font e software

Scarica tutte le opere

Proponente aggiuntivo

Scegliere nella parte di sinistra l'ordinamento:
Cronologico (per secolo, datazione incerta o elenco completo)
Alfabetico (per lettera iniziale o elenco completo)

Opere V secolo

- Adnotationes super Lucanum
- Adnotationes super Lucanum: supplementum
- Altercatio Hadriani Augusti et Epicteti philosophi
- Anecdotum medicum
- Ars (siue de nomine et uerbo)
- Ars Bobiensis [= Excerpta Bobiensia GL I 533-565]
- Ars Breuiata
- Ars de orthographia
- Ars grammatica (Asper minor)
- Ars grammatica (Cledonius)
- Ars grammatica (Ps. Asper)
- Brevis expositio Vergilii georgicorum
- Celeres passionnes
- Collatio Alexandri et Dindimi
- Commenta Bernensia in Lucanum
- Commentaria in metra Terentiana
- Commentarii in diuinationem in Quintum Caecilium
- Commentarii in Verrinas
- Commentarium de oratione et de octo partibus orationis artis secundae Donati
- Commentarius in Vergilii Aeneidos libros

1 2 3 4 5 6 »

Figura 3. Elenco delle opere del V secolo a disposizione dell'utente

Oltre a questi testi è poi possibile scaricare una serie di opere di critica sul tardoantico di cui si è prodotta un'edizione anastatica digitale in formato PDF e ricercabile per parola: si tratta di saggi liberi dal vincolo di copyright, oppure per cui si è ottenuto il consenso alla distribuzione da parte dei titolari.⁶

⁶ Punto di partenza sono state le opere di Italo Lana, tra i primi latinisti italiani a porre speciale attenzione al tardo antico e nel cui magistero si sono formati molti dei membri del progetto: un buon numero di questi studi, nati come dispense universitarie, non solo oggi possono essere reperiti con difficoltà, ma la loro stessa sopravvivenza è messa in pericolo dal rapido processo di disgregazione della carta acida su cui sono stati stampati.

1.2 La ricerca

All'utente della biblioteca si aprono due porte di accesso alla ricerca: una semplice e una avanzata. Nel primo caso è possibile interrogare il corpus mediante il semplice inserimento di una parola nella home page; le forme flesse si ottengono inserendo la radice della parola seguita da *. La marcatura a cui sono stati sottoposti i testi⁷ consente però anche interrogazioni più raffinate, possibili grazie alla maschera di ricerca avanzata: sono qui possibili ricerche non solo sull'universo dei testi, ma anche su sottosezioni stabilite dall'utente (opere singole, tipologie testuali specifiche, per secolo, ecc.). Asterischi e operatori booleani permettono di combinare le ricerche di più parole con desinenze variabili; determinabile è anche la prossimità delle parole oggetto di ricerca e la sezione testuale specifica in cui sono contenute (nell'intero testo, nei titoli, nei versi o nei passi in lingua greca).

The screenshot displays the digilibLT search interface. At the top, the logo 'digilibLT' is followed by the text 'Biblioteca digitale di testi latini tardoantichi' and logos for 'UPO' (Università del Piemonte Orientale) and 'REGIONE PIEMONTE'. A navigation bar includes links for 'home', 'il progetto', 'Notizie', 'Tardoantico nel web', 'Aiuto', 'Contatti e feedback', and 'Accedi'. The main search area is titled 'Cerca' and features a search input field with a 'CERCA' button. Below this, there are links for 'Opere disponibili per la ricerca', 'Ricerca avanzata', and 'Ricerca bibliografica'. A sidebar on the left offers 'Letture e download di testi' with radio buttons for 'Opere', 'Autori', and 'Bibliografia'. It also provides sorting options: 'Per ordine cronologico' (II, III, IV, V, VI, VII) and 'Per ordine alfabetico' (A-Z). A 'Risorse' section lists 'Opere sul tardoantico', 'Canone autori tardoantichi', 'Download di font e software', and 'Download testi'. The main search area is titled 'Ricerca all'interno dei testi' and includes a search input field, a 'Cerca' button, and an 'Azzera' button. It offers search logic options: 'tutte (AND)' (selected) and 'almeno una di (OR) queste parole'. A note states: 'Racchiudi le parole tra apici ("") per cercare sequenze esatte'. There is a field for 'Parole da escludere:'. Proximity options include 'parola(e)' (selected). Sectional search options are 'tutto il testo' (selected), 'titoli', 'verso', and 'passo in greco', with an 'Editor greco' button. The 'Criteri di selezione delle opere' section includes filters for 'Titolo', 'Autore', and 'Tipologia testuale', each with a search input and radio buttons for 'tutte' or 'almeno una di queste parole'. A 'Datazione' section has checkboxes for 'II', 'III', 'IV', 'V', 'VI', 'VII', and 'Datazione incerta'. A footer area contains the text 'Proponente aggiuntivo' and a circular logo.

Figura 4. Ricerca all'interno dei testi

7 Per una descrizione dettagliata delle scelte di marcatura del progetto cf. *infra* §4, con ulteriore bibliografia.

I risultati della ricerca danno accesso alle occorrenze dei termini cercati nelle singole opere, ciascuna voce permette poi l'accesso al testo completo dell'opera e a una concordanza (KWIC, Key Word In Context), visualizzabile seconda l'ordine del testo, per contesto sinistro o per contesto destro. L'immagine che segue (fig. 5) mostra la concordanza, in ordine del testo, delle ricorrenze di *pax* nell'*Epitoma historiarum Philippicarum Pompei Trogi* di Giustino.

The screenshot shows the DigilibLT website interface. At the top, there is a navigation bar with links like 'Home', 'Il progetto', 'Notizie', 'Tardoantico nel web', 'Aiuto', 'Contatti e feedback', and 'Accedi'. The main content area displays the search results for 'pax' in the text of 'Iustinus, Epitoma historiarum Philippicarum Pompei Trogi'. The results are organized by chapter and section, with the word 'pax' highlighted in red in the original image. The search results are as follows:

Chapter/Section	Text	Word	Text
II. 5.8	...partim suspensio vitam finierunt. Post haec	pax	apud Scythas fuit usque tempora...
V. 10.8	...ad ceteros iubet. Quibus rebus cum	pax	statuta esset, interiectis diebus repente...
VI. 6.5	...eo tantum insignis fuit, quod repente	pax	tota Graecia facta est, sed etiam eo...
VIII. 4.2	...misit; ibique ex commodo utrorumque	pax	facta. Ex ceteris quoque Graeciae...
XXII. 2.7	...potentiae recuperaret. Igitur non	pax	tantum Agathocli conciliatur, verum etiam...
XXXVII. 2.10	...annos X cum Seleuco pacem facit; sed	pax	ab hoste data interpellatur a fratre...
XXX. 3.10	...eo indutiae duorum mensium datae, ut	pax	, quae in Macedonia non conveniebat,...
XXXVIII. 4.1	...consulere, bellum sit cum Romanis an	pax	habenda; quin vero sit resistendum...

The interface also includes a search bar, a list of available works, and navigation options like 'Return to Search Risultati' and 'Vista stampabile'.

Figura 5. KWIC *pax* in Iustinus, *Epitoma historiarum Philippicarum Pompei Trogi*

1.3 Il contesto di supporto

La definizione di 'biblioteca' (*Library*) di DigilibLT sta a indicare una natura che non vuole limitarsi al database, ma rappresentare un ambiente di lavoro dedicato al tardo antico, all'interno del quale anche il semplice interessato possa trovare un orientamento entro un panorama letterario relativamente poco noto. Per questo motivo i testi si presentano accompagnati da schede che ne illustrano i contenuti e i principali problemi critici (scheda opera) così come i dati noti sull'autore (scheda autore), un apparato spesso a firma dei maggiori esperti dei testi in questione.

The screenshot shows the digilibLT website interface. At the top, there are logos for digilibLT (Biblioteca digitale di testi latini tardoantichi), UPO (UNIVERSITÀ DEL PIEMONTE ORIENTALE), and REGIONE PIEMONTE. A navigation bar contains links like Home, Il progetto, Notizie, Tardoantico nel web, Aiuto, Contatti e feedback, and Accedi. The main content area is titled 'De comoedia uel de fabula' by 'Euanthius' (saec. IV). It features a search bar, a 'CERCA' button, and a 'Scheda Autore' button. Below the title, there are download options for TXT, TEI, PDF, and E-PUB, along with a 'SCHEDA CATALOGRAFICA' button. The main text block contains a detailed introduction and analysis of the work, mentioning its relationship to Plautus and other Roman comedy forms. A 'Bibliografia' section on the right lists eight references. At the bottom, there are 'Operazioni preliminari' (Laura Mosca), 'Preparazione testo' (Beatrice Strona), and 'Codifica TEI' (Nadia Rosso). A 'Torna all'elenco' button is located at the bottom center.

Figura 6. Scheda di Giovanni Polara del *De comoedia vel de fabula* di Evanzio

La biblioteca comprende anche una ricca bibliografia, collocata sia a lato delle pagine che contengono le schede, sia accessibile dalla home page, dove un pulsante conduce a una maschera di ricerca avanzata.⁸ La figura 7 mostra il risultato di una ricerca su *Eumenius* in ordine alfabetico per autore moderno.

⁸ L'interrogazione è possibile attraverso diverse chiavi: autore moderno, autore antico, titolo, opera antica, tipologia editoriale (edizione, lessico, miscellanea, monografia), titolo di rivista, parole chiave descrittive del contenuto, anno di edizione, cronologia dell'autore e dell'opera.

The screenshot shows the digilibLT website interface. At the top, the logo 'digilibLT' is displayed next to the text 'Biblioteca digitale di testi latini tardoantichi'. To the right are logos for 'UPO' and 'REGIONE PIEMONTE'. Below the header is a navigation bar with links: Home, Il progetto, Notizie, Tardoantico nel web, Aiuto, Contatti e feedback, Accedi. On the far right of the navigation bar are icons for accessibility (A A A) and flags for Italy and the UK.

The main content area is titled 'Ricerca bibliografica'. It includes a search bar with a 'CERCA' button. Below the search bar, there are links for 'Opere disponibili per la ricerca', 'Ricerca avanzata', and 'Ricerca bibliografica'. A sidebar on the left contains sections for 'Lettura e download di testi' (with options for Opere, Autori, Bibliografia), 'Per ordine cronologico' (with options II, III, IV, V, VI, VII, Datazione incerta, Elenco autori), 'Per ordine alfabetico' (with options A, B, C, D, E, F, G, H, I, L, M, N, O, P, Q, R, S, T, V, Elenco autori), and 'Risorse' (with options Opere sul tardoantico, Canone autori tardoantichi, Download di font e software, Scarica tutte le opere). At the bottom of the sidebar is a 'Proponente aggiuntivo' section with a circular logo.

The search results are displayed in a list format. The results are sorted by 'Autore moderno' and are ordered alphabetically by author. The results are as follows:

- 1. Camastra P. Letteratura Latina Tardoantica. *Bibliografia riferita all'autore antico:*
 1. Eumenius saec. III
- 2. Cavallo G., Fedeli P., Giardina A. Lo Spazio Letterario di Roma Antica. Volume II: la circolazione del testo. *Bibliografia riferita all'autore antico:*
 1. Eumenius saec. III
- 3. Giardina A. - Silvestrini M. Il principe e il testo. *Bibliografia riferita all'autore antico:*
 1. Eumenius saec. III
- 4. Lana I. e Maltese E. V. Storia della civiltà letteraria greca e latina. Volume terzo: dall'età degli Antonini alla fine del mondo antico. *Bibliografia riferita all'autore antico:*
 1. Eumenius saec. III

Figura 7. Risultato di una ricerca su *Eumenius* in ordine alfabetico per autore moderno

2 Canone

Scelta preliminare per la creazione di una biblioteca digitale è la definizione del canone delle opere in essa contenute, con la relativa scelta dell'edizione di riferimento.

Sono necessarie alcune precisazioni. La biblioteca *digilibLT* comprende i testi in prosa dal II al VI secolo d.C.; è utile ribadire che siamo partiti da zero, senza un corpus digitalizzato pregresso (con parziale eccezione per quanto riguarda i grammatici su cui torneremo). Ciò ha implicato, in primo luogo, una selezione delle tipologie di testi da includere, in modo che il progetto fosse sostenibile per le forze a disposizione. Si è quindi rinviato l'inserimento di testi giuridici e cristiani, tenendo conto che per queste tipologie testuali esistono già risorse digitali (cf. *supra* nota 2).⁹

⁹ Come si è illustrato nel § 1; cf. anche Tabacco 2014, 3-6.

Inoltre non era disponibile un canone di riferimento esauriente né la sua composizione si è rivelata un'operazione semplice. Non vi sono comprese, infatti, solo opere di alto valore letterario, si pensi alla grande storiografia di Ammiano Marcellino, ma anche breviaristica, storia romanizzata, testi tecnici, manuali scolastici, *scholia*, commenti, testi geografici, antiquari, ecc., un vasto insieme di opere eterogenee e non facilmente classificabili. Solo alcune di esse erano già disponibili online, ma non inserite in una biblioteca esauriente e corredata di strumenti che permettessero una ricerca avanzata. Spesso, poi, i testi reperibili non garantivano correttezza e scientificità o non dichiaravano neppure l'edizione critica di riferimento. E infine, ed è un punto di non secondaria importanza, i siti esistenti, pur consentendo la libera consultazione unita a funzionalità di ricerca più o meno avanzate, generalmente non prevedono di offrire il download dell'opera completa.

Possiamo quindi affermare che la biblioteca *digilibLT* è stata creata ex novo per quanto riguarda sia la scelta del materiale da inserirvi sia la preparazione di testi digitalizzati.

Per quanto concettualmente la definizione del canone preceda la digitalizzazione delle opere, in realtà si è proceduto di pari passo, scansionando, correggendo e marcando i testi man mano che venivano messe a punto le varie voci del canone. Né sono mancate ridefinizioni delle voci del canone e sostituzioni delle edizioni di riferimento per adeguarsi, ad esempio, alle pubblicazioni scientifiche avvenute dal 2010, anno di inizio del progetto, a oggi. Ma procediamo con ordine.

La biblioteca *digilibLT* presenta in realtà un doppio canone. Online è disponibile l'elenco, consultabile per ordine cronologico o alfabetico, degli autori e delle opere. Tale elenco è stato preparato da Andrea Balbo e si è visto (§ 2.1) come le singole voci consentano di accedere alla lettura dei testi e delle relative schede opera e autore. Un consistente nucleo di tale elenco era già disponibile alla messa online della biblioteca, ma nel corso degli anni è stato continuamente aggiornato e ampliato, in base al progredire degli studi e parallelamente alla preparazione del canone vero e proprio, liberamente scaricabile in formato PDF, nel quale per ogni opera è indicata l'edizione di riferimento.¹⁰

Punto di partenza per l'individuazione sia delle opere sia dell'edizione di riferimento, è stato l'*Index* del *Thesaurus linguae Latinae*,¹¹ opera fondamentale per chiunque si accinga allo studio di testi dell'antichità latina. E tuttavia il *Thesaurus* fornisce un quadro di tutte le opere latine, rendendo

¹⁰ Il canone è stato preparato da Simona Musso, con aggiornamenti successivi a cura del gruppo di ricerca.

¹¹ La parte del volume *Praemonenda de rationibus et usu operis* (pubblicato nel 1990) dedicata all'indice è dovuta a D. Krömer.

complessa la consultazione a chi sia interessato in maniera specifica alle opere profane tarde, né è frequentemente aggiornato.¹² Per queste ragioni il canone realizzato nel progetto, oltre a svolgere la funzione strumentale di indicare le edizioni adottate, si presenta come utile strumento di lavoro per studiosi di tarda antichità. Rispetto all'*Index* del *Thesaurus* non si è operato solo un criterio selettivo, ma sono anche stati apportati aggiornamenti e modifiche, per i quali sono state determinanti, oltre agli studi condotti, le segnalazioni di studiosi di ogni parte del mondo che hanno permesso di rimediare a inevitabili imprecisioni e omissioni. In alcuni casi, ad esempio, sono state indicate opere che secondo la datazione canonica non rientrerebbero nel periodo preso in considerazione dal progetto, ma che sono state recentemente retrodatate¹³ oppure opere la cui redazione a noi giunta è di età medievale, ma che contengono nuclei di probabile datazione tardoantica.¹⁴ Altre volte, invece, avevamo ommesso opere rimaste inedite fino ad anni molto recenti o per le quali esistono solo edizioni risalenti al XVI secolo.¹⁵

La scelta dell'edizione di riferimento si basa sul criterio scientifico dell'individuazione della migliore esistente. Il panorama molto variegato delle tipologie testuali incluse a canone produce situazioni molto diverse rispetto alla disponibilità di edizioni. Per molte opere esiste una sola edizione, altre sono state edite solo in epoca prescientifica, mentre le opere più studiate sono leggibili in molte edizioni, alcune anche molto recenti, e non sono rari i casi in cui si possa usufruire di più edizioni di alto valore. Si è quindi operata una scelta che non intende, tuttavia, essere esclusiva. In una prospettiva futura, sarà possibile aggiungere l'apparato critico ai testi digitalizzati oppure affiancare più edizioni. Nel frattempo si è tenuto conto delle varie edizioni esistenti procedendo a un confronto costante, anzitutto per correggere eventuali errori di stampa: per citare un caso, si veda il *Breviarium* di Festo per il quale l'edizione di riferimento è Eadie (1967), ma si è tenuto conto delle correzioni tipografiche apportate nell'edizione Fele (2009).

Discorso a parte meritano i grammatici, per i quali il progetto si avvale dei testi che vennero digitalizzati all'Università di Torino sotto la direzione scientifica di Nino Marinone. Il prototipo di CD-Rom, che rappresentò l'esito del progetto (GL-CD), tuttavia, non venne mai commercializzato, e dopo poco più di due decenni mostra già la sua obsolescenza rendendo

12 Risalgono al 2010 gli *Addenda* all'*Index* dalla Bayerische Akademie der Wissenschaften e sono disponibili online (<http://www.thesaurus.badw.de/addenda.htm>).

13 Il progresso degli studi può aver modificato non solo la datazione, ma anche l'attribuzione di un'opera a un determinato autore.

14 Un caso esemplificativo è rappresentato dagli scolii.

15 Si pensi, ad esempio, alla versione latina dell'opera medica di Alexander Trallianus.

assai complessa l'estrazione dei dati in esso contenuti e la loro riconversione per uniformarli agli standard oggi in uso e garantirne l'accessibilità. Il *GL-CD* non utilizzava, infatti, la codifica Unicode UTF-8 per i caratteri e aveva adottato un font specifico per il greco (Greco), creato ad hoc (cf. Marinone 1979; Lomanto 1980). Ai fini dell'inserimento dei testi di *GL-CD* all'interno di *DigilibLT*, a essi è poi stata applicata una marcatura *TEI/XML*, coerente con le scelte della biblioteca, in modo da consentire una ricerca avanzata sul corpus.¹⁶

Non è ora necessario entrare nello specifico della codifica dei grammatici, ma il caso in questione ci offre il destro per sottolineare l'importanza della scelta del tipo di codifica dei testi al fine di garantirne la sopravvivenza. Il ritmo vorticoso di evoluzione degli strumenti digitali rende obsolete in pochissimi anni tecnologie che oggi ci paiono all'avanguardia. L'ambizione di progetti di digitalizzazione di testi è invece quella di durare nel tempo, ma ciò è possibile solo se si svincola il documento da hardware e software specifici o da rigide visualizzazioni predefinite, si adottano standard condivisi e si distribuiscono i testi sotto licenza *Creative Commons* e non protetti e limitati da copyright.¹⁷

Limitandoci alla trattazione dei grammatici relativamente al canone, bisogna spiegare che, partendo da un corpus pregresso, si è scelto inizialmente di adottare le stesse edizioni utilizzate nel *GL-CD*, procedendo poi, laddove esistano edizioni più recenti, ad aggiornare il corpus con l'edizione di riferimento attuale.¹⁸ La suddivisione delle opere e i relativi titoli adottati nel *GL-CD*, tuttavia, spesso non coincidono con quelli adottati dall'*Index* del *Thesaurus*, soprattutto per quanto riguarda i trattatelli anonimi e i frammenti. Problema accentuato, inoltre, dall'ambiguità dei titoli generici delle opere grammaticali (*Ars*, *De metris*, *De uerbo*, ecc., per fare solo alcuni esempi).

Per rendere univoca l'identificazione dei testi, è stato quindi necessario indicare a canone non solo l'edizione di riferimento (posta in grassetto), ma anche il volume e le pagine dell'edizione Keil, aggiungendo inoltre segnalazione di eventuali edizioni più recenti rispetto a quella adottata. Portiamo ad esempio il caso degli *Exempla Elocutionum* di Arusiano Messio.

16 La conversione in Unicode e la marcatura XML sono state operate da Piero Colombo-Step Srl. Data la frequente presenza di greco e di simboli tecnici (si pensi ai simboli metrici), al fine di garantire la massima correttezza, Ermanno Malaspina sta provvedendo a un controllo linguistico, correggendo manualmente eventuali inesattezze di transcodifica. Simona Musso, invece, si sta occupando della verifica della marcatura *TEI/XML*.

17 Per le licenze *Creative Commons* si veda il sito <http://creativecommons.org/>.

18 L'elenco delle opere del *GL-CD* e delle relative edizioni è contenuto in Lomanto 1980. Per maggiori dettagli sugli studi pubblicati a partire dal 1979 sul progetto di digitalizzazione dei grammatici cf. Tabacco 2014, 5 n. 8 e Borgna 2017, 4 n. 8.

Autore	Opera	Edizione di riferimento
Arusianus Messius saec. IV ex.	<i>Exempla elocutionum</i>	<p>Arusiani Messii <i>Exempla elocutionum</i>, introduzione, testo critico e note a cura di A. Di Stefano, Hildesheim 2011 (Collectanea Grammatica Latina 6).</p> <p>&</p> <p>Arusianus Messius, <i>Exempla elocutionum</i>, a cura di A. Della Casa, Milano 1977</p> <p>Cfr GL VII 449-514 <i>Grammatici Latini, VII. Scriptores de orthographia. Terentius Scaurus, Vellius Longus etc., ex recensione H. Keilii, Hildesheim 1961, 449-514 (reprografischer Nachdruck der Ausgabe Leipzig 1880).</i></p>

Figura 8. Voce del canone

Nella colonna di destra si vede in grassetto l'edizione di riferimento (Di Stefano 2011). Seguono l'indicazione dell'edizione digitalizzata nel GL-CD (Della Casa 1977) e il riferimento al volume e alle pagine dell'edizione Keil.

3 Formati e visualizzazione

L'indicazione dell'edizione di riferimento adottata non si trova solo nel canone, ma è riportata sia nella scheda opera online sia nel TEI Header dei testi digitalizzati, nel marcatore <sourceDesc> contenuto in <fileDesc>, come mostra l'esempio sotto riportato.¹⁹

19 I file TEI/XML sono suddivisi in due sezioni: il TEI Header - la sezione introduttiva in cui sono contenute informazioni relative all'opera, all'autore, alla fonte, alle modifiche apportate al testo - e il testo vero e proprio. I file del progetto *digilibLT* hanno il TEI Header suddiviso in quattro sezioni di cui l'ultima è opzionale. Nella prima sezione (<fileDesc>) sono indicati il titolo e l'autore dell'opera antica; il nome del responsabile della correzione linguistica e della marcatura (spesso non coincidenti); editore, luogo e data di pubblicazione; licenza *Creative Commons* sotto cui sono distribuiti i testi; il numero identificativo dell'opera e la citazione bibliografica dell'edizione di riferimento. La seconda sezione, <encodingDesc>, è a sua volta suddivisa in <projectDesc>, <editorialDecl>, <charDecl>. La parte più corposa è l'<editorialDecl> in cui si specificano sia le convenzioni seguite nella trascrizione e nella marcatura dei testi - distinguendo quelle generali da quelle specifiche di una singola opera - sia eventuali modifiche apportate al testo di riferimento. Il <projectDesc> riporta il nome esteso del progetto e dei suoi responsabili. Parleremo del <charDecl> affrontando la questione dei glifi. La terza sezione <profileDesc> contiene il <langUsage> in cui è specificata la lingua in cui sono scritti i testi del progetto *digilibLT*, cioè il latino. Il greco non è presente in tutti i testi e, laddove compare, costituisce generalmente una porzione nettamente minoritaria. È infine prevista la sezione <revisionDesc> in cui si può tener memoria di eventuali interventi sul file in tempi successivi alla prima pubblicazione. Per una descrizione più approfondita, cf. Musso 2014, 171-5.

```
<sourceDesc>
<p>
<hi rend="italic">Le premier mythographe du Vatican</hi>, texte étaabli par Nevio Zorzetti et traduit par Jacques Berlioz, Paris 1995 (Collection des Universités de France)
</p>
</sourceDesc>
```

Attraverso fogli di stile XSLT, dall'XML si ricavano i vari formati (HTML, .txt, ePub, PDF)²⁰ in ognuno dei quali compare la dichiarazione dell'edizione di riferimento. La scelta di *digilibLT* di rendere libero e gratuito il download dei testi, per intero²¹ e con la possibilità di scelta del formato più adatto alle proprie esigenze, rende la biblioteca unica nel suo genere.

Bisogna precisare che la scelta di ricavare formati con caratteristiche molto diverse fra loro da un unico file sorgente TEI/XML ha richiesto di riflettere con attenzione anche sulle possibilità di visualizzazioni. È ben noto che la marcatura TEI/XML ha un suo punto di forza proprio nel non essere vincolata a una specifica visualizzazione. La TEI, per quanto preveda anche marcatori indicanti semplicemente una formattazione,²² si caratterizza per marcatori descrittivi che determinano, ad esempio, il fatto che una porzione testuale sia un titolo, un'integrazione o una citazione e non come essa vada visualizzata. Si è già accennato che a partire dalla marcatura TEI si possono poi ottenere le più svariate visualizzazioni attraverso i fogli di stile, tuttavia ci sono alcuni casi specifici in cui determinate scelte di marcatura possono porre alcune limitazione nella visualizzazione.

Se ne può individuare un esempio nel caso in cui con il testo digitale si voglia riprodurre sia la scansione in libri, capitoli e paragrafi sia la suddivisione logica in capoversi dell'edizione di riferimento. I problemi nascono, infatti, non solo quando capitoli e paragrafi sono numerazioni indipendenti tra loro, ma anche quando uno dei due livelli strutturali (o a volte entrambi) non coincide con i capoversi, eventualità per nulla rare nei testi latini.

20 La natura dichiarativa e descrittiva di TEI/XML predica una caratteristica per una determinata porzione di testo (specificando, ad esempio, che si tratta di un titolo di paragrafo), ma non determina un output specifico per tale caratteristica (ad esempio, la sua visualizzazione in grassetto). Le opzioni di visualizzazione dei fenomeni marcati e i formati di output si ottengono attraverso fogli di stile XSLT (*Extensible Stylesheet Language-Transformation*); per la biblioteca *digilibLT* i fogli di stile sono stati realizzati da Fabio Ciotti. Per una descrizione dettagliata di TEI/XML si vedano il sito della TEI (<http://www.tei-c.org/>) dove sono disponibili le *Guidelines: TEI P5* (2015). Per un'introduzione più agile si veda Burnard, Sperberg-McQueen 2005, la cui edizione italiana è curata da Fabio Ciotti.

21 Si è fornita anche la possibilità di scaricare l'intero corpus in formato TEI/XML.

22 <hi rend="italic"> per il corsivo, <hi rend="bold"> per il grassetto e così via.

La marcatura TEI/XML è gerarchica e non permette la coesistenza di due strutture concorrenti di pari livello. Tuttavia il problema viene in parte superato con l'uso di marcatori vuoti che, invece di incorniciare la porzione testuale da descrivere, si limitano a essere situati nel punto di inizio di un determinato fenomeno.

I capitoli, ad esempio, sono marcati come `<div type="cap" n="">` all'inizio del capitolo e `</div>` alla fine. Se il capitolo contiene paragrafi, questi sono marcati come una `<div>` di livello inferiore `<div type="par" n="">` contenuto del paragrafo `</div>`. La marcatura di capitoli e paragrafi come `<div>` è molto comoda perché consente di scandire il testo in una struttura precisa, prevede per ogni `<div>` la possibilità di inserire titoli (`<head>`) e permette la creazione di alberi di navigazione funzionali alla lettura a schermo. Nei casi in cui, tuttavia, le numerazioni di capitoli e paragrafi non sono gerarchizzate, ma indipendenti e intrecciate, la marcatura con le `<div>` diventa impossibile:

```
<div type="cap" n="1">
<div type="par" n="1">testo del primo paragrafo</div>
<div type="par" n="2">testo del secondo paragrafo</div>
<div type="par" n="3"><p> inizio del testo del terzo paragrafo
</div>
<div type="cap" n="2">
proseguimento del testo del terzo paragrafo</p></div>
<div type="par" n="4"><p>testo del quarto paragrafo </p></div>
</div>
```

Per ovviare a tale problema, come si è detto, sono previsti marcatori puntuali. Ad esempio la biblioteca *digilibLT*, nei casi in cui si presenta conflitto, marca i paragrafi come `<milestone unit="par" n=""/>`. E, come si è accennato, il conflitto si genera non solo quando capitoli e paragrafi si incrociano, ma anche quando essi non coincidono con i capoversi.

La marcatura in `<div>` prevede, infatti, che il testo sia inserito non direttamente fra il tag di apertura e di chiusura delle `<div>` relative, ma all'interno del tag che delimita il capoverso `<p>` e che implica a capo (`<div><p> testo </p></div>`); ne consegue che il cambio di `<div>` determina l'inserimento di un capoverso. Il tag `<milestone/>`, invece, non implica l'inserimento di un capoverso, ma rispetto alle `<div>` è più limitante sia per l'approfondimento della marcatura sia per la creazione di alberi di navigazione sia per la possibilità di isolare con precisione la porzione testuale a cui fa riferimento.

Se nel caso di conflitto tra capitoli e paragrafi la scelta del `<milestone/>` al posto della `<div>` è obbligata, a meno di rinunciare a marcare una delle due numerazioni, nel caso in cui vi sia solo mancata coincidenza tra paragrafi/capitoli) e capoversi, vien da domandarsi se valga la pena di

rinunciare alle potenzialità della marcatura con <div> per non inserire degli a capo in punti non previsti dall'edizione di riferimento e spesso non in coincidenza con segni di interpunzione forte.

In realtà non esiste una risposta valida in assoluto; molto dipende, infatti, dalla finalità per cui si sono marcati i testi e dal prodotto finale che si vuole ottenere. Se i testi confluiscono semplicemente in una banca dati interrogabile, il fatto che vi sia una corretta scansione in capoversi diventa di secondaria importanza. Se invece si prevede di ricavare dall'XML anche formati in cui l'impaginazione ha un ruolo fondamentale, perché sono pensati per la lettura e la stampa e non solo per essere interrogati dalla macchina, allora la prospettiva è diversa. La scelta di voler ricavare PDF e ePub dall'XML ha implicato una scelta di marcatori che non ponesse limitazioni alle esigenze di visualizzazione. In tali formati, infatti, sarebbe fuori luogo un testo che andasse a capo a metà frase solo perché in quel punto vi è un cambio di paragrafo.

Nel .txt, invece, sarebbe irrilevante sia perché il formato non ha come utilizzo primario la lettura sia perché, essendo un formato editabile, può essere agevolmente modificato. Anch'esso, tuttavia, non è esente da problemi connessi alla visualizzazione, per quanto essi siano di tutt'altro genere.

Come è noto questo formato non consente di conservare le formattazioni, ma tale mancanza avrebbe comportato una perdita di significati non trascurabile. Le edizioni di riferimento dei testi della biblioteca *digilibLT* sono eterogenee e usano formattazioni e diacritici con i più svariati significati. Abbiamo provveduto a normalizzare i diacritici più comuni e le formattazioni di più semplice interpretazione,²³ ma non è stato possibile uniformare completamente tutti i testi perché sarebbe stato necessario uno studio approfondito per ognuno di essi, e ciò avrebbe allungato di molto i tempi di pubblicazione.²⁴ Si è quindi deciso di utilizzare i marcatori

23 Ad esempio l'uso del grassetto o del corsivo per indicare i titoli di capitolo o paragrafo. In questi casi si è adottata la marcatura <head> e attraverso i fogli di stile i titoli si visualizzano in grassetto. Una sintesi delle normalizzazioni effettuate è contenuta nella nota al testo inclusa in TEI Header nell'<editorialDecl> e leggibile nei frontespizi dei formati PDF e ePub, nonché nella pagina online dedicata alle informazioni editoriali di ogni testo. Chi desiderasse maggiori dettagli può inoltre consultare le *Note di trascrizione e di codifica* (scaricabili dal sito all'indirizzo http://www.digiliblt.uniupo.it/opere_tardoantico/Note_di_trascrizione_e_di_codifica.pdf) e Musso 2014, 172-5.

24 Una distinzione di tutti gli usi del corsivo (integrazione, integrazione solo in presenza di lacuna materiale, correzione del testo dei codici, citazione, titolo di opera, nomi di autori, ecc.) o delle altre tipologie di formattazione, con relativa marcatura specifica, avrebbe rallentato il lavoro senza garantire una soluzione ottimale a ogni esigenza e avrebbe impedito di conseguire lo scopo principale del progetto: mettere a disposizione degli studiosi, in tempi brevi, un vasto corpus di testi liberamente ricercabile e scaricabile. Approfondimenti di marcatura sono sempre possibili in una seconda fase. Vengono comunque specificati per ogni testo eventuali usi particolari della formattazione.

previsti dalla TEI per descrivere le formattazioni: `<hi rend="">`; il valore dell'attributo `@rend` indica il tipo di formattazione (italic, bold, ecc.).

Ma anche qualora avessimo attribuito una marcatura descrittiva specifica a ogni fenomeno, in fase di visualizzazione si sarebbe comunque dovuto ricorrere alle formattazioni per indicare che quella porzione ha un valore peculiare. Ne consegue, quindi, che ogni informazione rappresentata dalla formattazione sarebbe andata persa nel .txt.

Per evitare tale impoverimento, si è deciso di utilizzare una sequenza di caratteri speciali come indicatori delle diverse formattazioni (si veda la tabella 1).

Formattazione	Marcatura	Segnalazione in .txt
Corpo minore	<code><hi rend="small">testo</hi></code>	<code>{@testo@}</code>
Corsivo	<code><hi rend="italic">testo</hi></code>	<code>{€testo€}</code>
Grassetto	<code><hi rend="bold">testo</hi></code>	<code>{\$testo\$}</code>
Sottolineato	<code><hi rend="underline">testo</hi></code>	<code>{#testo#}</code>
Esteso	<code><hi rend="expanded">testo</hi></code>	<code>{%testo%}</code>

Tabella 1. Indicatori delle diverse formattazioni

Tale prospetto si trova in testa ai file .txt, affinché l'utente sia informato di ogni modifica apportata ai testi presenti nella biblioteca.

4 Glifi

Riflessioni sulle modalità di visualizzazione sono state necessarie anche per rappresentare al meglio quei simboli non ancora presenti nello standard Unicode o, se presenti, non inclusi nei font più comuni. Per quanto molto ricco, Unicode inevitabilmente non può prevedere e comprendere ogni glifo esistente. Tali simboli sono particolarmente numerosi nei testi in prosa tardoantica, data la natura tecnica di molti di essi, e per lo più si possono raggruppare in due tipologie: numerali con moltiplicatori e unità di misura, per i quali si sono adottate due soluzioni distinte.²⁵

Per i numerali si è scelto di esplicitarne il valore numerico. A testo si scrive il numerale senza moltiplicatore e con la marcatura si dichiara

²⁵ Inizialmente si era pensato di adottare un font gratuito creato appositamente per studiosi di testi antichi (Cardo) e di segnalare agli utenti della biblioteca *digilibLT* il sito da cui è possibile scaricarlo. Tale opzione, però, ha presto mostrato inconvenienti. Anzitutto neppure Cardo è in grado di soddisfare ogni esigenza, dato che non rappresenta tutti i simboli necessari. Inoltre, i glifi non presenti in Unicode sono stati inseriti nell'area ad uso privato comportando un'errata visualizzazione non solo a chi non ha scaricato la versione più aggiornata di Cardo, ma anche nel formato TXT ricavato dall'XML. Si è quindi dovuta cercare una soluzione alternativa utilizzando la marcatura prevista dalla TEI.

come deve essere interpretato. Per rappresentare, ad esempio, il simbolo indicante 10000, X con linea orizzontale soprascritta, si scrive:

```
<num value="10000">X</num>
```

la cui visualizzazione è:

X(10000)

Nel caso di unità di misura, e in generale per i simboli che non rappresentano numeri, è necessaria una marcatura più articolata. In TEI Header si descrivono i vari caratteri fornendo a ciascuno un identificatore univoco.

```
<encodingDesc>
<charDecl>
<char xml:id="char1">
<charName>Nome del simbolo</charName>
<desc>Descrizione dell'aspetto del simbolo</desc>
</char>
<char xml:id="char2">
<charName>Nome del simbolo</charName>
<desc>Descrizione dell'aspetto del simbolo</desc>
</char>
</charDecl>
</encodingDesc>
```

A testo si rinvia al carattere in questione col tag vuoto <g/>. Qualora il simbolo esista, anche se solo nell'area a uso privato di un font specialistico, esso viene inserito all'interno dell'elemento <g>:

```
<text>
<body>
<div>
<p>testo <g ref="#char1">—</g> testo. </p>
<p>testo <g ref="#char2"/> testo. </p>
</div>
</body>
</text>
```

Attraverso i fogli di stile si è stabilito che accanto al simbolo, o al suo posto, comparisse tra parentesi il nome dell'unità di misura, per ognuna delle quali si è stabilita un'abbreviazione, il cui scioglimento è riportato nel marcatore <editorialDecl>.

Den. per Denarius
 Drag. per Dragma
 Pon. per Pondus
 Scrip. per Scripulus
 Sem. per Semis
 Semunc. per Semuncia
 Sext. per Sextarius
 Vict. per Victoriatus
 Vnc. per Vncia
 ecc.

Vediamo un esempio sia di unità di misura per cui Unicode (o un font nell'area a uso privato) prevede il simbolo sia di unità per cui il simbolo è assente.

In TEI Header, nel <charDecl>, si esplicitano gli identificatori dei simboli, per ognuno dei quali si aggiungono l'abbreviazione del nome e la descrizione del suo aspetto.

```
<charDecl>
<char xml:id="char1">
<charName>Vnc.</charName>
<desc>Tractus longus</desc>
</char>
<char xml:id="char2">
<charName>Sext.</charName>
<desc>S cum linea transuersa</desc>
</char>
</charDecl>
```

A testo, nel marcatore <g>, si richiama l'identificativo del glifo in questione:

```
<milestone unit="par" n="30"/> Anethi surculi uruntur eiusque pulueris
cribrati <g ref="#char1">—</g> I, cicharbae <g ref="#char1">—</
g> III remittes in olei optimi <g ref="#char2"/> I et adicies gallae
tunsae semunciam inlinesque eo porriginosum et ulcerosum capite
contra pilum et postea lauabis in balneo.
```

Il testo riportato ha la seguente visualizzazione:

[30] Anethi surculi uruntur eiusque pulueris cribrati —[Vnc.] I, cicharbae —[Vnc.] III remittes in olei optimi [Sext.] I et adicies gallae tunsae semunciam inlinesque eo porriginosum et ulcerosum capite contra pilum et postea lauabis in balneo.

Tale marcatura da un lato supplisce a una carenza di Unicode e dall'altra permette di attuare una schedatura dei simboli presenti nelle edizioni,²⁶ operazione necessaria per poter chiedere a Unicode di ampliare lo standard.

5 Natura collaborativa e qualità: un rapporto problematico

Le questioni prese in esame mostrano le scelte adottate dalla biblioteca *digilibLT* di fronte alle problematiche sollevate dal processo di digitalizzazione. Va da sé che si tratta di una delle varie possibilità offerte dalla TEI, che non impone uno schema rigido e univoco di marcatura. Anzi, la sua natura aperta permette anche di introdurre nuovi marcatori purché siano dichiarati attraverso DTD (*Document Type Definition*) o XSD (*XML Schema*). Se tale versatilità permette di adattare al meglio la marcatura alle esigenze di chi digitalizza, è pur vero che non favorisce la compatibilità fra testi realizzati in progetti diversi, che pur utilizzano il vocabolario TEI. Non mancano tentativi di creare vocabolari condivisi per testi di tipologie specifiche, come dimostra *Epidoc* per i documenti epigrafici antichi,²⁷ tuttavia si è ancora lontani dalla possibilità di creare un unico database testuale interrogabile nel suo complesso partendo dai testi latini presenti nelle varie biblioteche digitali liberamente consultabili online.

Questo non significa, tuttavia, che non vi possano essere proficui rapporti di collaborazione fra progetti diversi o fra studiosi.²⁸ La biblioteca *digilibLT* ha infatti creato una rete di contatti con specialisti del settore, non solo italiani, che hanno fornito il proprio contributo a vari livelli.

È necessario chiarire le modalità in cui avviene il rapporto di collaborazione. Il web odierno tende sempre più a favorire l'inserimento diretto di dati da parte dell'utente, ma il gruppo di ricerca ha preferito esercitare un controllo accurato delle informazioni presenti nella biblioteca prima della loro immissione per garantire non solo, ovviamente, la correttezza dei dati, ma anche la coerenza e l'uniformità nelle modalità di preparazione dei dati.

Abbiamo già menzionato l'importanza delle segnalazioni ricevute nella compilazione del canone (§ 3) e l'arricchimento per la biblioteca prodotto da schede realizzate dai massimi specialisti del testo in questione (§ 2.3).

26 Tale schedatura è stata avviata da David Paniagua, esperto di testi tecnici e progressivamente ampliata dai collaboratori della biblioteca *digilibLT* via via che si sono riscontrati nuovi simboli e glifi. La complessità di tale schedatura è accentuata dal fatto che le varie edizioni critiche utilizzano glifi diversi per rappresentare la stessa unità di misura. Riportiamo a titolo esemplificativo, tra parentesi quadre, alcune rappresentazioni dell'oncia [— · / (~)].

27 Le *Guidelines* sono reperibili all'indirizzo <http://www.stoa.org/epidoc/gl/latest/>.

28 Il sito della biblioteca prevede una pagina specifica di contatti e feedback (<http://www.digiliblt.uniupo.it/contatti.php>).

In alcuni casi, inoltre, il testo digitalizzato è stato preparato da chi ne sta curando l'edizione critica, facendo sì che l'edizione digitale presenti un testo criticamente più accurato delle edizioni cartacee esistenti.²⁹

Infine sono preziosi i contributi e le segnalazioni ricevuti per eliminare refusi rimasti nella preparazione dei testi. Garantire la massima correttezza dei testi è stato un obiettivo centrale del progetto. A tal scopo sono state valutate le varie procedure di digitalizzazione per adottare quella che garantisse la minor incidenza di errore. La biblioteca *digilibLT* ha consapevolmente scelto l'opzione controtendenza di mantenere all'interno del gruppo di ricerca tutte le fasi di digitalizzazione, compreso il controllo del testo da marcare, non demandando questa fase a ditte esterne. Si è già avuto modo di riflettere su tale questione,³⁰ in questa sede ci soffermeremo solo su alcuni aspetti. Si tenga presente, come si è detto, che la biblioteca è partita da zero, senza un corpus già digitalizzato (a eccezione dei grammatici). Gran parte del tempo e delle energie è quindi stata dedicata a tale operazione preliminare.

Per creare un testo digitale sostanzialmente vi sono due vie: o viene riscritto, o viene resa editabile l'immagine dell'edizione cartacea del testo dell'edizione di riferimento.³¹ La prima strada non era sostenibile per un gruppo di dimensioni ridotte, si è quindi optato per la seconda via, puntando a ottenere un'alta qualità delle immagini digitalizzate attraverso l'uso di scanner professionali *Atiz BookDrive*. Tali scanner sono caratterizzati da un piano a V su cui poggia il libro aperto e ognuno dei due piatti viene fotografato attraverso macchine reflex full frame (nello specifico si tratta di Canon 5D) che permettono di ottenere immagini ad altissima risoluzione. I software in dotazione (*BookDrive Capture* e *BookDrive Editor*) consentono non solo di regolare le impostazioni delle macchine in fase di scatto, ma anche di ottimizzare le immagini ottenute per migliorare la resa dell'OCR.³²

Nonostante questo, l'incidenza di errore dei programmi di riconoscimento ottico dei caratteri è rimasta elevata. Ciò si spiega non solo con i difetti meccanici della fonte cartacea, ma anche con le difficoltà linguistiche presentate dai testi. Anzitutto per il latino tardo non ci sono validi dizionari di supporto al programma di riconoscimento, senza contare che, laddove un errore generasse una parola esistente, nessun dizionario potrebbe se-

29 Si veda, ad esempio, il caso del commento di Servio all'*Eneide* di Virgilio, che *DigilibLT* (per gentile concessione dell'editore stesso) offre in anteprima nell'edizione di Giuseppe Ramires.

30 Cf. Musso 2014, 154-7 e per una riflessione più generale Lana 2012a, 76-83.

31 Togliendo ovviamente tutto ciò che non è testo antico: apparato, note, introduzione, commento, ecc.

32 Ad esempio aumentando il contrasto laddove vi siano pagine sbiadite o ingiallite.

gnalarlo. Le grafie non canoniche, l'uso del greco e di numerosi simboli, inoltre, contribuiscono a ostacolare un buon riconoscimento dei caratteri.

È emersa immediatamente la necessità di un approfondito controllo umano all'esito dell'OCR. Inizialmente si pensava di poter evitare una lettura integrale dei testi procedendo al confronto automatico di due distinti riconoscimenti della stessa opera.³³ Tale espediente funziona bene se il numero di errori è limitato ma, nel nostro caso, l'alta incidenza d'errore avrebbe costretto alla revisione di quasi tutte le righe di testo. Si è quindi optato per una duplice lettura integrale dei testi, affidata a studiosi e studenti che avessero una buona competenza in lingua latina. Quest'ultimo particolare ci ha permesso in numerosi casi di segnalare, ad esempio, errori tipografici presenti nell'edizione di riferimento.

Ovviamente siamo ben consci del fatto che nemmeno una duplice lettura può garantire l'individuazione di tutti i refusi. Ed è per questo motivo che assume grande importanza il controllo ulteriore che possono esercitare gli utenti sui testi online. Si sono rivelate di grande utilità le segnalazioni effettuate da un collaboratore del *Thesaurus linguae Latinae*, Johann Ramminger, che attraverso un'analisi automatica dei testi ha isolato una serie di forme sospette. In alcuni casi si trattava effettivamente di refusi o di errori già presenti nell'edizione di riferimento ma molte, in realtà, erano forme tarde e ciò ha messo in chiara evidenza il valore rappresentato dal corpus della biblioteca *digilibLT* per uno studio dell'evoluzione della lingua latina.

6 Digitale, perché?

Se, ad esempio, nel caso delle dispense universitarie degli anni '60, stampate su carta acida, il mezzo digitale rappresenta uno strumento di salvaguardia della memoria, abbiamo però spesso notato come la rapidissima evoluzione della tecnologia esponga i prodotti digitali della ricerca all'elevato rischio dell'obsolescenza dei supporti, che non sempre riescono ad adattarsi alle diverse versioni dei sistemi operativi. Vi sarebbe infatti bisogno di una manutenzione costante e a lungo termine, pratica virtuosa ma resa poco probabile dalla scadenza dei progetti o dalla scarsità di fondi. Nasce così un acceso dibattito all'interno del mondo accademico, che se da un lato riconosce il valore dello strumento digitale in termini di velocità di accesso e come vettore di rapida diffusione della cultura, non di rado guarda ancora alla *digital scholarship* con qualche perplessità,³⁴ relativa

33 Attraverso programmi specifici quali ad esempio WinMerge.

34 In merito, ancora valide le riflessioni di McGann 2005, 77, secondo cui i progetti digitali «typically get born into poverty - even the best-funded ones. Ensuring their maintenance,

non solo al problema della caducità dei supporti, ma anche a quello – assai più controverso – del controllo scientifico e del valore di quanto prodotto con metodologie digitali in varie sedi (cf. Tabacco 2014, 7-8; Musso 2014, 186-7).

La progettazione ex novo di *DigilibLT*, che non aveva alle spalle una banca dati in CD-ROM, neppure in forma di prototipo, ha altresì favorito una riflessione complessiva non solo in termini tecnici, ma anche ma anche sulle potenzialità del progetto ad ampio raggio. Si è così scelto, come già è stato detto, di mantenere all'interno del progetto tutti i passaggi della lavorazione dei testi, una decisione che ha quindi imposto l'allargamento del gruppo di ricerca. A questo fine *DigilibLT* ha guardato in modo particolare ai giovani, facendone tuttavia, non una mera forza lavoro, ma costruendo intorno a loro un vero e proprio percorso formativo in filologia digitale.³⁵

6.1 Un punto di partenza

Una banca dati online marcata secondo XML-TEI non rappresenta, tuttavia, solo un risultato raggiunto: essa deve anche costituire un punto di partenza per nuovi approcci e ricerche. Già nel 2003 John Unsworth tracciando un bilancio dello sviluppo fino ad allora delle *digital humanities* affermava:

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 answer.

Più di dieci anni dopo, queste domande rimangono attuali:

a tuttora, l'enfasi sembra essere stata posta prevalentemente sulla produzione, piuttosto che sull'impiego dei dati così prodotti, come se si trattasse soltanto di formare un immenso archivio di testi codificati,

development, and survival is a daunting challenge. Worst of all, the work regularly passes without much practical institutional notice. Accepted professional standards do not control the work in objective ways. Most of it comes into being without oversight or peer review». Si veda anche la lettera inviata dal direttivo dell'AIUCD (Associazione per l'Informatica Umanistica e la Cultura Digitale), Dino Buzzetti nel febbraio 2014 all'allora ministro dell'Istruzione, Università e Ricerca Stefania Giannini in merito alla questione della valorizzazione, anche in sedi concorsuali, di quanto prodotto con metodologie digitali (per il testo della lettera cf. <http://www.aiucd.it/osservazioni-critiche-dellaiucd-sullasn/>).

³⁵ Del valore formativo di *DigilibLT* si è già dato conto in Borgna 2017, 13-16.

predisposti per così dire a ‘futura memoria’ senza considerare più di tanto il problema dell’uso e delle modalità di impiego di tali risorse. (Buzzetti 2011, 39)

Il sostegno di *DigilibLT* all’*open access* come canale di disseminazione della ricerca, una scelta obbligata laddove tale ricerca sia sostenuta da finanziamenti pubblici, è così in grado di dare origine ad altre iniziative, nella volontà di rifuggire quell’isolamento a cui è spesso soggetta la *Digital Scholarship*.³⁶ In questo senso la biblioteca si è posta come base di partenza per il progetto TBL, *Textual Bilingualism in Latin* che indaga i fenomeni di contatto linguistico tra greco e latino nei testi letterari latini tardi.³⁷ L’esperienza maturata e la strumentazione acquisita ha poi portato alla fondazione, presso l’Università di Torino (proponente aggiuntivo di *DigilibLT*), del Centro di Ricerca Interdipartimentale per la digitalizzazione e la realizzazione di Biblioteche Digitali Umanistiche MEDIHUM (*Memoria Digitalis Humanistica*).³⁸ Da ricordare anche GeoLat, Geografia per la letteratura latina, che in vista di un accesso ai testi della latinità tarda a partire da mappe, ha elaborato e recentemente diffuso Go!, la prima ontologia geografica esplicitamente studiata per il mondo antico.³⁹

In conclusione, alla domanda sempre più pressante di istituire un circolo virtuoso tra risorse letterarie online, ontologie e *linked data*, come grande opportunità per lo sviluppo futuro non solo delle *Digital Humanities*, ma degli studi umanistici in genere (cf. Ciotti 2014, 9), *DigilibLT* sta quindi rispondendo non solo con la sua stessa costante crescita, ma anche ponendosi come motore di sviluppo per altre iniziative, segno che l’investimento sui giovani e l’attenzione per l’aspetto collaborativo e l’accesso aperto (mai privi, tuttavia, di un rigoroso controllo scientifico) si stanno rivelando intuizioni corrette.

36 «Digital scholarship - even the best of it - is all more or less atomized, growing like so many Topsyies. Worse, these creatures are idiosyncratically designed and so can’t easily talk to each other» (McGann 2005, 77).

37 Cf. <http://tbl.uniupo.it> (direzione scientifica: Maria Napoli)

38 Cf. <http://www.medihum.unito.it>. Il centro coniuga esigenze di usabilità e accesso comune ai contenuti culturali e di ricerca, digitalizzati o nativi digitali, con l’orizzonte scientifico delle attività di text e data mining e di linguistica computazionale per l’estrazione e l’analisi dei dati testuali

39 Cf. <https://goo.gl/3VRPGt> (direzione scientifica: Maurizio Lana). In merito si vedano anche Ciotti et al. 2016.

Bibliografia

- Borgna, Alice (2016): «Latino e informatica. Una proposta di et... et nella società dell'aut... aut». *ClassicoContemporaneo*, 3 (2017), 1-22.
- Burnard, Lou; Sperberg-McQueen, Michael (2005). *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari* [online]. A cura di Fabio Ciotti. Milano: Sylvestre Bonnard. URL http://www.tei-c.org/Vault/P4/Lite/teiu5_it.html (2017-10-17).
- Buzzetti, Dino (2011). «Oltre il rappresentare. Le potenzialità del *markup*». Fiermonte, Perilli 2011, 39-62.
- Ciotti, Fabio (2014). «Digital Literary and Cultural Studies. State of the Art and Perspectives». *Between*, 8(4), 1-17.
- Ciotti, Fabio et al. (2016). «Ontologies and the Cultural Heritage. The case of GO!» [online]. Draelants, Isabelle et. al. (eds.), *Proceedings of the II International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016)*, 7-18. URL <http://ceur-ws.org/Vol-1595/> (2017-11-10).
- Eadie, John William (1967). *The Breviarium of Festus. A Critical Edition with Historical Commentary*. London: Athlone Press.
- Fele, Maria Luisa (2009). *Il breviarium di Rufio Festo*. Hildesheim: Weidmann.
- Fioromonte, Domenico; Perilli, Lorenzo (a cura di) (2011). *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*. Firenze: Le Lettere.
- Guidelines: TEI P5 (2015). «Guidelines for Electronic Text Encoding and Interchange» [online]. URL <http://www.tei-c.org/Guidelines/> (2017-10-17).
- Lana, Maurizio (2011). «Un *database* testuale per il latino tardo». Fioromonte, Perilli 2011, 281-300.
- Lana, Maurizio (2012a). *Biblioteche digitali. Un'introduzione*. Bologna: Bononia University Press.
- Lana, Maurizio (2012b). «Da una *digital library* del latino tardo ad un corpus globale». Ciotti, Fabio; Crupi, Gianfranco (a cura di), *Dall'informatica umanistica alle culture digitali = Atti del convegno di studi in memoria di Giuseppe Gigliozzi* (Roma, 27-28 ottobre 2011). Roma: Digilab, 134-50.
- Lana, Maurizio (2012c). «Metodologie e problematiche per una biblioteca digitale. Il caso di *digilibLT*». *DigItalia. Rivista del digitale nei beni culturali*, 7, 40-64.
- Lomanto, Valeria (1980). «Lessici latini e lessicografia automatica». *Memorie dell'Accademia delle Scienze di Torino. Classe di Scienze Morali, Storiche e Filologiche*, V(4), 113-270.
- Marinone, Nino (1979). «Concordanza dei Grammatici Latini». Suppl. di *Atti dell'Accademia delle Scienze di Torino. Classe di Scienze Morali, Storiche e Filologiche*, 112, 9-19.

- McGann, John J. (2005). «Culture and Technology. The Way We Live Now, What Is to Be Done?». *New Literary History*, 36(1), 71-82.
- Mosca, Laura; Musso, Simona (2016). «Costruire una biblioteca digitale di testi latini tardoantichi. Problemi e scelte». Maconi, Ludovica; Marazzini, Claudio (a cura di), *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori = Atti del Convegno "Piazza delle Lingue", VIII Edizione* (Firenze 6-7 novembre 2014). Firenze: Edizioni dell'Accademia della Crusca, 143-58.
- Musso, Simona (2014). «La marcatura di testi latini tardoantichi. Un compromesso ragionevole». *Aevum Antiquum*, 11, 151-87.
- Tabacco, Raffaella (2000). *Itinerarium Alexandri*. Testo, apparato critico, introduzione, traduzione e commento a cura di Raffaella Tabacco. Firenze: Leo S. Olschki Editore.
- Tabacco, Raffaella (2014). «La tarda antichità latina tra i codici e il web». *Aevum Antiquum*, 11, 3-17.
- Tabacco, Raffaella (2016). «La prosa latina pagana tardoantica e la biblioteca digitale *digilibLT*». *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori = Atti del Convegno "Piazza delle Lingue", VIII Edizione* (Firenze 6-7 novembre 2014). Firenze: Edizioni dell'Accademia della Crusca, 125-41.
- Unsworth, John (2003). «Tool-Time, or 'Haven't We Been Here Before?'. Ten Year in Humanities Computing» [online]. *Transforming Disciplines. The Humanities and Computer Science* (18th January 2003). Washington DC. URL <http://people.virginia.edu/~jmu2m/carnegie-ninch.03.html> (2017-11-10).

Epigrafia digitale o digitalizzazione delle epigrafi?

Flavia De Rubeis

(Università Ca' Foscari Venezia, Italia)

Abstract At present, the issue of digital epigraphy seems limited to the digitalization of epigraphs by means of the creation of databases. Digital epigraphy, unlike the digital palaeography that in the last few years has known a potential development that is likely to produce very interesting results, still does not have its own defined search line, at least at current research.

Keywords Medieval palaeography. Medieval epigraphy. Digital humanities. Digitalization. Database.

As of today, there are numerous projects concerned with developing such methodologies. These encompass a wide range of scientific, interdisciplinary approaches such as forensic document analysis, optical character recognition, quantification of “scribal fingerprints”, metric analysis, quantitative methods, advanced manuscript analyses such as DNA and imaging techniques such as multi-spectral digitisation, classification systems and databases. Although some achievements have been made already, much research is still required. For instance, something as seemingly fundamental as the automated recognition of characters in handwritten texts has proven extremely complex, due largely to the very wide variation in styles of handwriting, the often poor quality of surviving manuscripts, the lack of standard orthographies which complicates prediction, and so on. Such computational methods as proposed by digital palaeography have been the subject of research in the last few years, but most of this has been theoretical or applied only to small cases, partly because of the very high degree of labour that is typically involved. The applications to date have also focused almost exclusively on the question of scribal identity, ignoring other aspects of palaeography. Furthermore, they tend to view letter-forms as objects outside the manuscript or documentary context in which they were written, but palaeographers have long understood that handwriting depends heavily on the context in which it is produced. (Hassner et al. 2013)

Così Hassner, Rehbein, Stokes e Wolf introducono il lavoro sulla paleografia digitale e i problemi, le difficoltà e i vantaggi di un approccio digitale alla paleografia.

In un recente incontro, tenutosi a marzo del 2016, la questione della paleografia viene vista sotto un differente profilo, con un approccio meno problematico:

Medievalists already have an impressive track record in using the Digital Humanities in their teaching and research. It is now a common practice for an undergraduate to scrutinize high-resolution images of a thousand-year-old manuscript while she transcribes the texts it preserves. A researcher can pin point when and where certain words were used, or identify the scribe who copied a particular manuscript, or map out the paths of literary borrowings thanks to online reference tools.¹

Come si può osservare, l'approccio alla digitalizzazione dei materiali paleografici è visto sotto due prospettive decisamente differenti che possono essere considerate come rappresentative di due differenti approcci alla paleografia digitale: un primo, che pone il problema del riconoscimento delle forme grafiche al centro della ricerca; un secondo che identifica nella possibilità di trascrivere un testo manoscritto il punto focale delle cosiddette *Digital Humanities* rivolte al manoscritto e ai testi da questi trasmessi.

Spostando su di un differente ambito la riflessione e senza entrare qui nel merito della discussione che interessa la cosiddetta paleografia digitale, tema sul quale si discute da lungo tempo, vorrei soffermarmi sui testi traditi attraverso le epigrafi, ossia passare a una differente classe di materiali scrittori, le iscrizioni, ossia le scritture esposte.

Una iscrizione, o una scrittura esposta, per sua natura reca un testo destinato alla pubblica esposizione, sia essa per ampie o per ridotte fasce di potenziali lettori. La specificità della volontà 'pubblicistica' che determina la realizzazione dell'iscrizione, ossia la volontà della committenza a rendere una determinata azione documentabile e pubblica, nonché duratura il più a lungo nel tempo, rende inoltre il manufatto strettamente legato ai contesti per i quali o nei quali è stato realizzato o conservato. Ne consegue che i manufatti epigrafici sono per loro stessa natura direttamente legati con il luogo di messa in opera principale: affreschi, iscrizioni lapidee, lignee o metalliche, su vetro o tessuti. E, ancora in conseguenza di questo stretto e intimo rapporto tra iscrizione e luogo di destinazione o di conservazione, la sola garanzia per la sopravvivenza del manufatto è il mantenimento della sua funzione di uso. Così, ad esempio, perché una iscrizione funeraria locativa possa svolgere la sua funzione, sarà necessario che la sepoltura alla quale è legata non sia violata o distrutta o altro. Una volta venuta meno la funzione di uso, una iscrizione, quale che sia il committente o il testo con-

¹ Call-for-Papers: *Digital Britain: New Approaches to the Early Middle Ages* (March 25-26, 2016). Cf. il sito web <https://digitalbritain.weebly.com/> (2017-10-18).

servato o il prestigio posseduto al momento della sua realizzazione, ebbene questa iscrizione è destinata invariabilmente al degrado.

Si aggiunga poi a questo anche il luogo di conservazione o di messa in opera principale: una epigrafe esposta sulla facciata di una chiesa, ammesso che nel frattempo non siano intervenuti fattori di danneggiamento volontari o accidentali, difficilmente sfuggirà ai danneggiamenti da cause naturali. Il degrado prodotto dall'azione delavante dell'acqua, le concrezioni calcaree che si depositano sulle superfici possono provocare perdita parziale o integrale di testi epigrafici.

Un testo può andare perduto anche solo parzialmente, o con degrado dei caratteri, quali ad esempio alterazioni chimiche dei pigmenti su affreschi, per i quali solo di recente si parla di recuperi da restauro legati all'uso di nanotecnologie e in grado di contrastare le alterazioni degli affreschi stessi e di restituirne quindi la leggibilità, non solo in termini iconografici ma anche, dove presenti, con restituzioni testuali.

A questi problemi, già importanti per le conseguenze che essi comportano (perdita e scomparsa di testi), si aggiungano anche le difficoltà generate da questa particolare classe di manufatti, ossia la lavorazione della scrittura: scritte a solchi, a rilievo, alveolate, in metallo, a niello, *pictae*, pre- o post- consolidamento degli intonaci o dell'intonachino. Qui i problemi sono molteplici e sono legati al supporto scrittoria, alla morfologia delle lettere, alla tecnica di lavorazione delle epigrafi, allo stato di conservazione del manufatto e infine allo stato di conservazione del testo.

Come si è visto in precedenza, uno dei problemi che interessa la paleografia digitale o l'epigrafia digitale, risiede nell'esigenza di creare uno strumento in grado di elaborare la paleografia digitale, ossia di arrivare a stabilire forme e variazioni delle forme con riconoscimento automatico di queste. Ora, se questo è possibile per classi di scritte che presentano come supporto materiale una medesima materia scrittoria (papiro, pergamena, carta), e mi riferisco in particolare alla scrittura in ambito latino, diverso è il caso delle iscrizioni.

Questa diversità di supporti scrittori, tra manufatti librari e manufatti epigrafici ha determinato anche lo sviluppo di differenti percorsi di ricerca per la paleografia digitale: da un lato, con il manoscritto librario, si cerca di analizzare non solo il testo, ma anche di creare una paleografia digitale in grado di riconoscere in automatico le forme scritte; questo affianca gli strumenti per la trascrizione dei testi, dove però il problema deve essere posto su due fronti differenti, ossia la trascrizione automatica delle lettere e il riconoscimento automatico delle lettere (con la sua paleografia).

Per l'ambito strettamente epigrafico, la digitalizzazione dei materiali epigrafici porta alla creazione di database, talvolta con la creazione di archivi epigrafici digitali interattivi. Ma pur essendo archivi digitali interattivi, per i quali si prevedono le interrogazioni semplici o complesse, più

raramente è possibile interagire con il testo e il suo supporto materiale, sebbene esistano database con tali caratteristiche.

La digitalizzazione delle epigrafi, in generale, passa quindi per la trasposizione di un testo epigrafico in una banca dati in cui, con eventuali approfondimenti relativi al supporto materiale, sono presenti le trascrizioni dei testi con le eventuali edizioni; spesso queste banche dati, specialmente se originate da precedenti pubblicazioni (come si vedrà qui di seguito) analogiche, sono corredate da riproduzioni fotografiche. Così trattati i materiali appaiono come repertori online di materiali epigrafici già trattati in pubblicazioni sul cartaceo: in altre parole, quello che si potrebbe definire come libro informatico.

I vantaggi di questi repertori sono innegabili: immagini digitalizzate (non sempre di qualità eccelsa, soprattutto se recuperate da un preesistente cartaceo); reperibilità di testi; possibilità di aggiornamento bibliografico (sempre che i siti siano seguiti e aggiornati); sistemi di interrogazioni trasversali; trasposizione di materiali librari dalla base analogica alla forma digitale; link tra repertori epigrafici.

Fin qui però non si intravede ancora, per l'ambito epigrafico, il salto di qualità che al contrario ha interessato la produzione scrittoria libraria. Per questi ultimi, come bene è stato sottolineato da Peter Stokes e il suo gruppo nella citazione che ho richiamato in apertura del presente lavoro, le difficoltà risiedono proprio nella natura del testo manoscritto, autografo: la variabilità dei supporti, la variabilità della scrittura e del singolo scrivente. A questo poi si aggiungono le difficoltà rilevate dal gruppo di ricercatori, ossia la tendenza a sviluppare progetti su micro aree della scrittura, quali la mano di un singolo copista, o un gruppo di manoscritti aventi punti di contatto comuni che li rendono identificabili come gruppo a sé stante e via dicendo. Ora, se queste difficoltà sono state espresse nei confronti dei materiali librari, non sono contemplate le scritture epigrafiche per le quali i progetti di digitalizzazione al momento sono limitati alla realizzazione di archivi digitali epigrafici e non al superamento di questa fase in favore di un qualcosa di differente, come potrebbe essere la possibilità di rendere maggiormente leggibili epigrafi attraverso il rilievo digitale e quindi la creazione di una banca dati interattiva con immagini mobili, bi- o tridimensionali.

In tal modo le banche dati esistenti, come ho già richiamato in precedenza, presentano alcune limitazioni e tali limiti credo che possano essere sintetizzati come segue:

- non esistono archivi digitali epigrafici interattivi analoghi ad archivi digitali dei libri manoscritti (quindi con la possibilità di 'leggere' un'edizione direttamente sul manufatto);
- esiste, ma non è posto, il problema della paleografia epigrafica (tenendo conto che allo stato attuale degli studi, la paleografia epigra-

fica ancora non è pienamente sviluppata come lo è invece quella dei materiali librari e documentari);

- esiste il problema dei supporti materiali e della loro riproduzione (una fotografia frontale potrebbe non permettere ad esempio di studiare la rigatura di un manufatto o di accentuare effetti chiaroscurali non esistenti), poiché la maggior parte delle riproduzioni fotografiche digitali sono statiche, non bi- o tri-dimensionali; dal momento che la maggior parte delle banche dati parte da una precedente versione analogica dei materiali epigrafici, di frequente gli archivi epigrafici digitali sono costituiti dalla semplice trasposizione in formato digitale dei materiali stessi, con le loro note di edizione e commenti di varia natura;
- esistono poi problemi di natura 'epigrafica', richiamati in precedenza e legati alla natura dei supporti scrittori e allo stato di conservazione dei testi che costituiscono una variabile non irrilevante per l'edizione critica epigrafica.

In particolare, per una corretta edizione e valutazione di un'iscrizione, l'attenzione deve essere rivolta anche all'aspetto materiale del manufatto: aspetti quali la rigatura, le tecniche di scrittura, la natura del supporto materiale, lo stato di conservazione, la classe di appartenenza dell'oggetto recante il testo epigrafico e, non da ultimo, la paleografia del manufatto stesso. Questi aspetti trovano normalmente la loro sede naturale nell'apparato di commento all'edizione epigrafica. Si tenga presente, tuttavia, che proprio questi aspetti sono i più statici nell'apparato descrittivo di un testo epigrafico digitalizzato. Infatti, mentre una edizione critica può aprire a varianti, differenti lezioni o interpretazioni, secondo parametri che appaiono più vicini alla ecdotica di un testo che non alla digitalizzazione di un manufatto epigrafico, gli aspetti materiali, quando presenti, sono descrittivi e non sono reperibili attraverso una riproduzione attenta anche a questi aspetti specifici del manufatto epigrafico.

In tal modo, quando si parla di digitalizzazione delle epigrafi, si intende la digitalizzazione epigrafica o digitalizzazione delle epigrafi? Nella maggior parte dei casi, dagli archivi digitali disponibili in rete, si ha a che vedere con digitalizzazione delle epigrafi e non digitalizzazione epigrafica. Si tratta, in altre parole, di utilissimi strumenti di ricerca, dotati di sistemi di interrogazione efficaci (nella maggior parte dei casi), di frequente muniti di riproduzioni fotografiche, ma pur sempre di digitalizzazioni di epigrafi, nei migliori dei casi.

Diverso è invece il caso dell'epigrafia digitale. Per questa ultima penso che l'aspetto della riproduzione del manufatto vada portato in primo piano e che debba essere vista anche in un'ottica di approccio interattivo del manufatto e non della sola edizione chiusa. Ho già richiamato la necessità di verificare, attraverso le immagini, lo stato di salute dei manufatti. Questo non solo in relazione alla conservazione del manufatto, ma anche

in prospettiva di offrire la possibilità di leggere, attraverso le immagini, i testi specialmente quando gravemente compromessi da stato di conservazione precari.

In quest'ottica, sono stati sviluppati software mirati alla riproduzione digitale di manufatti, software che offrono la possibilità, mediante una serie di acquisizioni di immagini e quindi il trattamento dei dati attraverso specifici software, di effettuare veri e propri rilievi quasi tridimensionali, come ad esempio il RTI, Reflectance Transformation Imaging.²

Questo tipo di approccio, mirato all'oggetto e alla sua restituzione digitale, potrebbe offrire la possibilità di superare le difficoltà espresse in precedenza e dovute ai supporti materiali delle iscrizioni; potrebbe permettere valutazioni a distanza di manufatti, con la visualizzazione a luce orientata mobile dei manufatti e quindi potrebbe aprire, penso, la porta alla paleografia digitale epigrafica.

In questa direzione l'epigrafia digitale permetterebbe di contestualizzare la scrittura all'interno del manufatto, di indicarne le specificità e quindi di vedere il manufatto nella sua interezza, superando quella lontana, valida ma pur sempre limitante visione dei testi epigrafici intesi come puro testo senza contesto.

Bibliografia

Hassner, Tal; Rehbein, Malte; Stokes, Peter A.; Wolf, Lior (2013). «Computation and Palaeography. Potentials and Limits (Dagstuhl Perspectives Workshop 12382)» [online]. *Dagstuhl Manifestos*, 2(1). DOI 10.4230/DagMan.2.1.14.

2 URL <http://culturalheritageimaging.org/Technologies/RTI/> (2017-10-18).

Parte 2.

Strumenti di ricerca semantica
e per la critica testuale

Dal *Thesaurus* di BIA alle ontologie di BIA-Net Prospettive di sviluppo e integrazione con altre basi di conoscenze

Marco Giunti
(Università degli Studi di Cagliari, Italia)

Abstract In the first part of this paper, we will describe the essential lines of an ontology (implemented in OWL) for the three archives of the *Bibliotheca Iuris Antiqui* (BIA). Namely, we will briefly outline the classes and the relations of this ontology and will give a graphic representation of its conceptual structure. In the second part, we will discuss the perspectives of development that the ontological form of a knowledge basis opens up. It is pointed out that, on the one hand, the ontological form allows the development of radically new forms of information search and retrieval and, on the other one, it allows to link and integrate in a natural way different knowledge bases, each one characterised by specific criteria of internal organisation. In this regard, the essential steps that need to be taken to bring about such an integration, and the possible benefits that will ensue, will be analysed.

Sommario 1 Introduzione. – 2 Un'ontologia per i tre archivi di BIA. – 2.1 Schema ontologico del *Thesaurus*. – 2.2 Schema ontologico di *Opera*. – 2.3 Schema ontologico di *Fontes*. – 2.4 Relazioni fra *Thesaurus*, *Opera* e *Fontes*. – 3 Prospettive di sviluppo. – 3.1 La ricerca semantica e il paradosso di Menone. – 3.2 Connessione e integrazione con altre basi di conoscenza: Linked Open Data e Semantic Web.

Keywords Semantic Web. Linked Open Data. RDF. OWL. Semantic search. Meno's Paradox.

1 Introduzione

Il progetto BIA-Net è nato dall'esigenza di rendere accessibile sul web la base di conoscenze della *Bibliotheca Iuris Antiqui* (BIA). Quest'ultimo è un sistema informativo integrato sui diritti dell'antichità, già in precedenza digitalizzato, ma finora accessibile solo mediante CD-ROM (Palazzolo 2002). Il progetto, finanziato nell'ambito del PRIN 2009, ha visto la partecipazione di cinque unità (Università degli Studi di Catania, Università degli Studi di Perugia, Consiglio Nazionale delle Ricerche, Università 'Magna Graecia' di Catanzaro, Università degli Studi di Cagliari) e si è concluso nell'ottobre 2013 con l'implementazione di una prima versione di biblioteca digitale in rete, BIA-Net, che utilizza tecno-

logie open source e standard internazionali per la gestione degli archivi e dell'interfaccia web.

La base di conoscenze BIA è composta da tre parti fondamentali: *Opera*, un archivio bibliografico di oltre 50.000 opere che trattano il diritto romano e, più in generale, i diritti dell'antichità; *Thesaurus*, un sistema di classificazione dei documenti presenti in *Opera*, costituito da una gerarchia di alcune centinaia di classi e da circa 10.000 concetti che descrivono tali classi; e infine *Fontes*, l'archivio digitalizzato delle fonti dei diritti dell'antichità.

Nel quadro degli obiettivi generali del progetto, l'unità di Cagliari ha offerto il proprio supporto teorico a una revisione del *Thesaurus* di BIA (Palazzolo 1995), finalizzata anche allo studio di un sistema di reperimento automatico dell'informazione secondo rapporti di rilevanza. In questa prospettiva, è stato fondamentale esplicitare e definire l'ontologia sottostante non solo al *Thesaurus*, ma anche agli altri due archivi, *Opera* e *Fontes*, di cui BIA è composta. L'ontologia è stata implementata in *Web Ontology Language Description Logic* - OWL DL (W3C 2004a).

Nella prima parte di questo lavoro vengono descritte le linee essenziali di tale ontologia per i tre archivi di BIA. In particolare, ne sono brevemente descritte le classi e le relazioni e viene data una rappresentazione grafica della sua struttura concettuale.¹ Nella seconda parte, sono invece discusse le prospettive generali di sviluppo che si aprono una volta che una base di conoscenze sia posta in forma ontologica. In particolare, si pone in evidenza come la forma ontologica permetta, da un lato, di sviluppare modalità radicalmente nuove di ricerca e reperimento dell'informazione e, dall'altro, di collegare e integrare in modo naturale basi di conoscenze diverse e con specifici criteri di organizzazione interna. A questo proposito, sono chiariti quali siano i passi essenziali da compiere nella prospettiva di tale integrazione e i vantaggi che se ne potrebbero ricavare.

2 Un'ontologia per i tre archivi di BIA

BIA è una base di conoscenze costituita da tre parti interconnesse: *Opera*, *Thesaurus* e *Fontes*. *Opera* è un archivio bibliografico di oltre 50.000 opere che trattano del diritto romano e, più in generale, dei diritti dell'antichità. Nel *Thesaurus* si trovano delle classi (alcune centinaia) che permettono di classificare i contenuti dei documenti presenti in *Opera*, e dei concetti (circa 10.000) che descrivono tali classi. *Fontes* è l'archivio delle fonti dei diritti dell'antichità. Si dà sotto una sommaria descrizione delle classi e

¹ Una descrizione più dettagliata dello schema ontologico di BIA si trova in Giunti, Sergioli, Vivinet 2014.

della relazioni che costituiscono l'ontologia o, più propriamente, lo schema ontologico, di ciascuno dei tre archivi.²

Si tenga presente che una qualsiasi relazione il cui nome è del tipo 'X1aaaaX2... Xn' è stata implementata come una *object property* di OWL, e dunque ha la relazione inversa. Tuttavia, per semplicità, l'inversa non è indicata nei diagrammi. Tutte le altre relazioni sono *datatype properties* di OWL; i loro valori sono dunque letterali e nessuna di tali relazioni ha la relazione inversa.

2.1 Schema ontologico del *Thesaurus*

Il *Thesaurus* si presenta in due forme diverse ma completamente equivalenti: classificato o alfabetico. Facendo riferimento alla forma classificata, ciascuna classe del *Thesaurus* è identificata da un codice numerico. A ciascuna classe corrisponde una vedetta che può essere espressa in cinque lingue (italiano, inglese, tedesco, francese, spagnolo), oppure solo in latino. Per esempio, la vedetta della classe con codice '1' è espressa dai cinque termini: 'FONTI', 'SOURCES', 'QUELLEN', 'SOURCES', 'FUENTES', mentre la classe il cui codice è la stringa '1.1.1.1.1' ha la vedetta espressa dal solo termine latino 'LEGES'. Si noti che a classi diverse può corrispondere la stessa vedetta. Per esempio, le classi 1.1.1.1.1 e 1.1.2.5.1 hanno la stessa vedetta, LEGES.

Ciascun codice rappresenta biunivocamente una classe del *Thesaurus*. Le classi possono essere fra loro nella relazione di sottoclasse (in questo caso i relativi codici sono del tipo: '1'; '1.1'; '1.1.1'). Classi appartenenti allo stesso livello gerarchico hanno codici del tipo '1.1'; '1.2'; '1.3', e così via.

Inoltre, ciascuna classe può avere associati dei descrittori, che sono concetti semanticamente pertinenti alla stessa. Essi sono indicati nel *Thesaurus* classificato mediante termini latini. Per esempio, i descrittori della classe 1.1.1.3 (fig. 1) sono i dieci concetti corrispondenti ai seguenti termini latini o latinizzati: 'basilica', 'de actionibus', 'ecloga', 'epanagoge aucta', 'manuale legum', 'prochiron', 'synopsys basilicorum', 'synopsis minor', 'Theophili paraphrasis', 'Tipucitus'. Si noti che lo stesso concetto può essere descrittore di più classi. Per esempio, il concetto 'aurum' è descrittore di ambedue le classi 3.2.2 e 5.7.

A ciascun descrittore di una data classe può risultare associato un gruppo di altri concetti, espressi da termini latini. Essi, al solo fine della descrizione della classe in questione, sono sussunti sotto il relativo descrittore. Per esempio, i concetti *libri basilicorum* e *scholia basilicorum* sono sus-

² Le classi dell'ontologia non devono essere confuse con le classi del *Thesaurus* di BIA. In effetti, dal punto di vista dell'ontologia, le classi del *Thesaurus* sono a tutti gli effetti individui (istanze) dell'ontologia stessa.

sunti sotto il descrittore *basilica* della classe 1.1.1.3 (fig. 1). Il rapporto tra ciascuno di tali concetti, il descrittore a cui esso è associato e la classe descritta determina quindi una relazione ternaria, che nell'ontologia è denominata 'X1èSussuntoInX2èDescrittoreDiX3'.

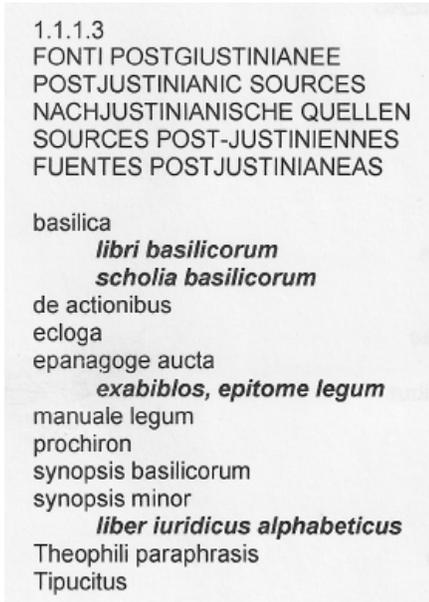


Figura 1. La classe 1.1.1.3 come appare nel *Thesaurus* classificato di BIA

Tra i concetti sussunti sotto un dato descrittore di una data classe può sussistere una relazione di equivalenza,³ che può esprimere sinonimie, varianti ortografiche, o simili. Tale equivalenza non è da intendersi in modo assoluto, ma soltanto relativamente al particolare descrittore sotto il quale i concetti sono sussunti e alla particolare classe che tale descrittore descrive. Per esempio, i concetti 'exabiblos' ed 'epitome legum' sono equivalenti relativamente a 'epanagoge aucta', che li sussume e che, a sua volta, è descrittore della classe 1.1.1.3 (Figura 1). Il rapporto tra i due concetti equivalenti, il descrittore che li sussume e la classe descritta determina quindi una relazione quaternaria,⁴ che nell'ontologia è denomi-

3 Nel *Thesaurus* classificato l'equivalenza fra due concetti è indicata da una virgola che separa i termini corrispondenti, scritti sulla stessa riga. Si vedano, per esempio, i termini 'exabiblos' e 'epitome legum' in figura 1.

4 Come noto, l'introduzione di relazioni a più di due posti in linguaggi quali RDF, RDFS (W3C 2004b) e OWL (W3C 2004a) non è possibile in modo diretto, in quanto tali linguaggi ammettono soltanto relazioni binarie (Corcho, Gómez-Pérez 2002). Questa limitazione è usualmente superata adottando particolari pattern ontologici (Gangemi, Presutti, 2009),

nata 'X1equivalenteX2èSussuntoInX3èDescrittoreDiX4'. Analoga relazione di equivalenza può sussistere anche tra un concetto e un descrittore di una classe. In tal caso, il rapporto tra il concetto equivalente, il descrittore e la classe descritta determina una relazione ternaria, che è denominata 'X1equivalenteX2èDescrittoreDiX3'. Tutte le classi e relazioni che costituiscono lo schema ontologico del *Thesaurus* sono mostrate nella figura 2.

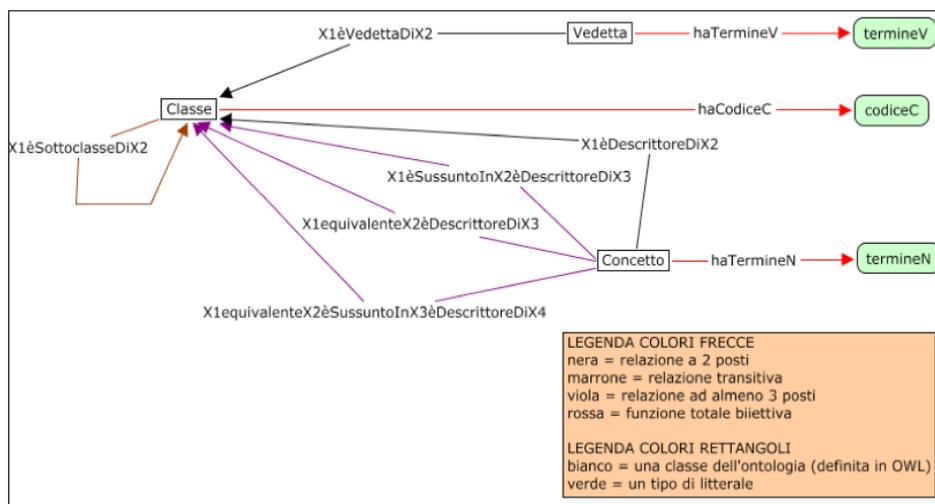


Figura 2. Schema dell'ontologia del *Thesaurus*

2.2 Schema ontologico di *Opera*

Per ogni opera, la relativa scheda bibliografica indica il titolo, l'autore, la pubblicazione (rivista, editore, atti di, ecc.), il numero delle pagine (per i libri) o i numeri della prima e ultima pagina (per gli articoli), l'anno di pubblicazione e, infine, il testo (che può essere o l'abstract dell'opera, se l'opera è un articolo, o l'indice, se l'opera è un libro). Le classi e relazioni dello schema ontologico di *Opera* sono mostrate nella figura 3.

che permettono di rappresentare una relazione n-aria ($n > 2$) mediante opportune relazioni binarie (W3C 2006). Il pattern ontologico utilizzato per la rappresentazione in OWL delle relazioni a tre e quattro posti del Thesaurus di BIA è particolarmente adeguato per tener conto dei rapporti di rilevanza semantica (Giunti 2010) tra le diverse entità presenti nell'ontologia. Tale pattern è oggetto di un lavoro attualmente sottoposto a revisione: Giunti, Marco; Sergioli, Giuseppe; Vivanet, Giuliano (2017). *La rappresentazione delle relazioni n-arie nel Semantic Web*.

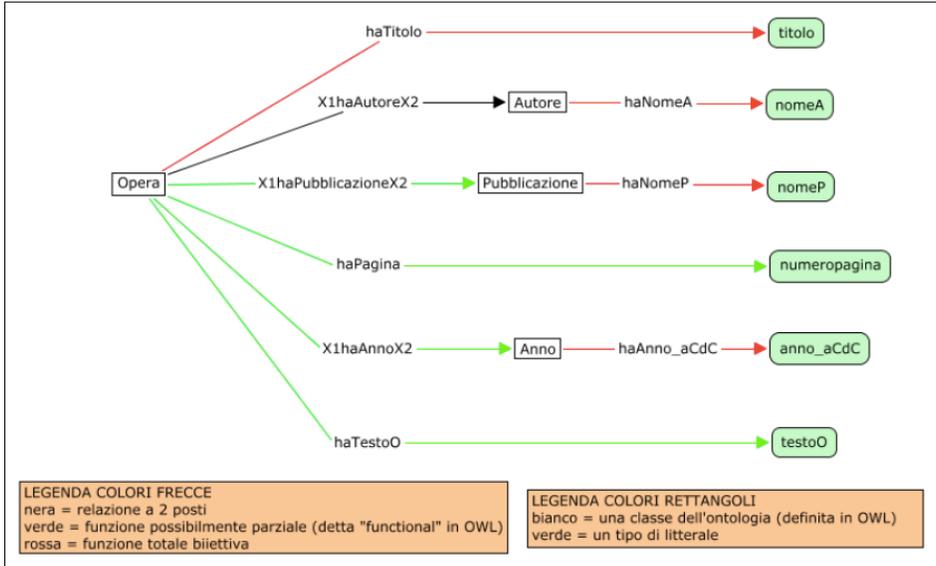
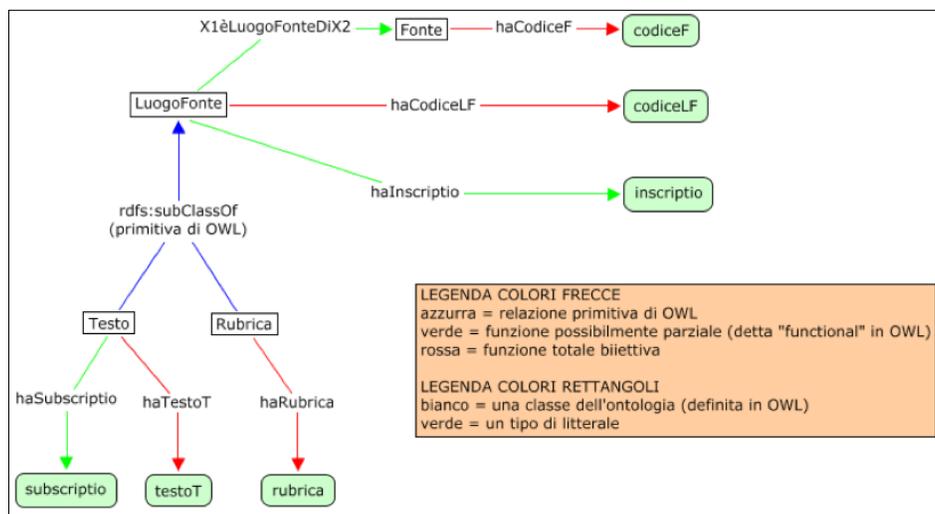


Figura 3. Schema dell'ontologia di *Opera*

2.3 Schema ontologico di *Fontes*

Ciascuna fonte è univocamente identificata da un codice. Inoltre, ciascuna fonte è suddivisa in parti elementari, dette luoghi fonte, che sono anch'essi univocamente identificati da codici. I luoghi fonte possono essere o di tipo testo (in tal caso potremmo trovare, oltre al testo vero e proprio, anche una *subscriptio*), o di tipo rubrica (in tal caso avremo il testo della rubrica). Inoltre, alcuni luoghi fonte (sia di tipo testo sia di tipo rubrica) possono contenere una *inscriptio*. Le classi e relazioni dello schema ontologico di *Fontes* sono mostrate nella fig. 4.

Figura 4. Schema dell'ontologia di *Fontes*

2.4 Relazioni fra *Thesaurus*, *Opera* e *Fontes*

Oltre alle relazioni interne a ciascuno dei tre archivi *Thesaurus*, *Opera* e *Fontes*, sono presenti anche relazioni che li interconnettono. Le relazioni fra *Thesaurus* e *Opera* sono due. La scheda bibliografica di una qualsiasi opera, infatti, indica una o più classi del *Thesaurus* a cui quell'opera appartiene e, di solito, indica anche uno o più descrittori di tali classi. Nel primo caso, si dirà che l'opera è classificata in una certa classe; nel secondo, che essa ha come termine controllato un certo concetto (ovvero, il descrittore indicato).

Fra *Opera* e *Fontes* c'è una sola relazione, anch'essa desumibile dalla schede bibliografiche. La scheda di un'opera, infatti, molto spesso indica uno o più luoghi fonte che sono analizzati o discussi nell'opera stessa.

Infine, non ci sono relazioni dirette fra *Thesaurus* e *Fontes* ma, ovviamente, ce ne sono due indirette: fra le classi o i concetti del *Thesaurus* e i luoghi fonte di *Fontes*, attraverso le opere di *Opera*. Tutte le relazioni fra i tre schemi ontologici sono mostrate nella fig. 5.

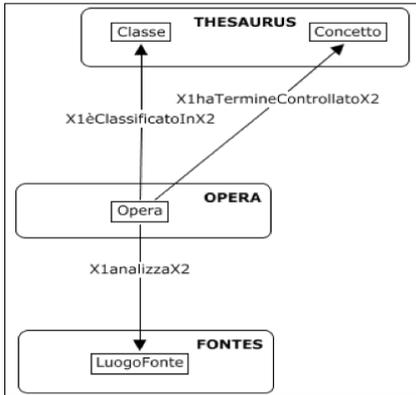


Figura 5. Relazioni fra i tre archivi di BIA

3 Prospettive di sviluppo

Per il momento abbiamo considerato soltanto lo schema dell'ontologia di BIA, cioè le classi e le relazioni che permettono di esprimere tutte le informazioni contenute nei tre archivi *Thesaurus*, *Opera* e *Fontes* nella forma di asserzioni (o triple) RDF. Per completare l'ontologia di BIA è quindi necessario procedere al suo popolamento, ovvero alla specificazione di tutte le istanze di ciascuna classe e all'asserzione delle relazioni in cui tali istanze si trovano. In questo modo, la base di conoscenze dei diritti dell'antichità, attualmente organizzata in forma tradizionale nei tre archivi distinti di BIA, sarà infine trasformata in un'unica rete semantica espressa nella forma di un grafo RDF.

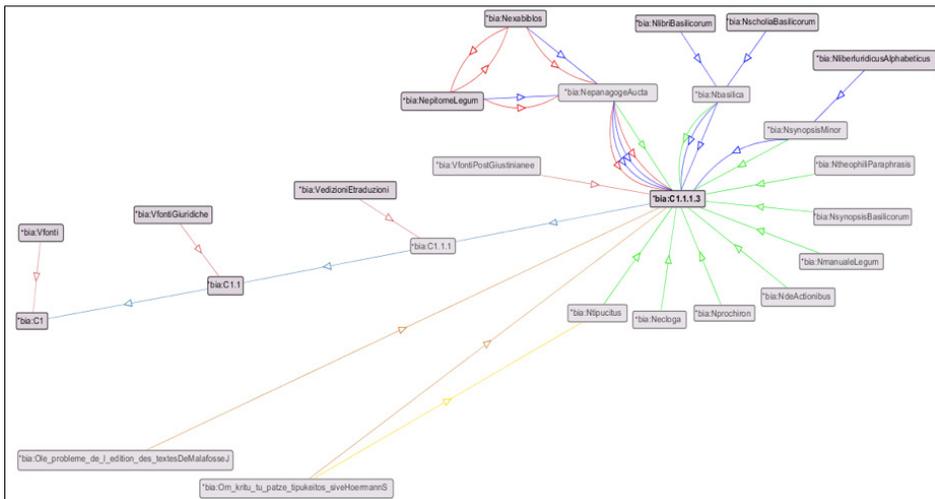


Figura 6. Le informazioni riguardanti la classe 1.1.1.3 espresse in forma ontologica

Per avere un'idea di che cosa significhi passare da una forma tradizionale di rappresentazione della conoscenza a quella ontologica, nella forma di grafo RDF, si confronti la fig. 1 con la fig. 6. Il grafo di quest'ultima esprime tutte le informazioni che, nella fig. 1, sono espresse in forma scritta. Ciò che appare subito evidente è che la forma ontologica esplicita le relazioni che sussistono fra le istanze di cui si tratta.⁵ Tali relazioni sono invece implicite, e quindi non immediatamente utilizzabili, nella rappresentazione tradizionale.

Più in dettaglio, le diverse relazioni sono rappresentate nel grafo in fig. 6 da frecce di diversi colori: $X1 \text{ èDescrittoreDi} X2$ —verde; $X1 \text{ èSussuntoIn} X2 \text{ èDescrittoreDi} X3$ —blu; $X1 \text{ equivalente} X2 \text{ èSussuntoIn} X3 \text{ èDescrittoreDi} X4$ —rosso; $X1 \text{ èVedettaDi} X2$ —ocra. Queste sono le relazioni necessarie per esplicitare tutta l'informazione relativa alla classe 1.1.1.3, come appare in forma scritta nel Thesaurus classificato (fig. 1).

È importante notare che nel grafo sono inoltre rappresentate anche altre informazioni presenti in BIA. Tali informazioni utilizzano le tre ulteriori relazioni: $X1 \text{ èSottoclasseDi} X2$ —celeste; $X1 \text{ èClassificatoIn} X2$ —marrone; $X1 \text{ haTermineControllato} X2$ —giallo. I due nodi (rettangoli) del grafo in basso a sinistra rappresentano due opere e non sono quindi istanze del *Thesaurus* ma piuttosto di *Opera*. Esse sono però connesse al *Thesaurus* in quanto la scheda bibliografica dell'opera più a sinistra specifica che è classificata in 1.1.1.3, mentre la scheda dell'altra opera indica che, oltre a essere classificata in 1.1.1.3, ha anche il termine controllato 'Tipucitus'. Infine, i quattro nodi collegati da frecce celesti rappresentano classi del *Thesaurus*, con le relative vedette (collegate a esse da frecce ocra). La classe più a destra è la 1.1.1.3; successivamente, da destra a sinistra, si trovano le classi 1.1.1, 1.1 e 1. Ciascuna di queste quattro classi è sottoclasse (freccia celeste) della successiva.

3.1 La ricerca semantica e il paradosso di Menone

Una volta completato il popolamento dell'ontologia, la struttura reticolare delle informazioni risulterà esplicita. Più precisamente, nella rete saranno contenuti due tipi di informazioni: (i) le istanze dell'ontologia, rappresentate dai nodi della rete e (ii) le relazioni asserite (o i fatti) relative a queste istanze, rappresentate da particolari percorsi della rete.

Siccome le informazioni saranno inserite in una rete, esse saranno utilizzabili da opportune applicazioni semantiche. Infatti, in primo luogo, è possibile definire una misura generale della rilevanza di un'informazione i rispetto a

⁵ Nell'esempio in questione (fig. 6) le istanze sono: la classe 1.1.1.3, tutti i concetti che la descrivono e la sua vedetta. Tali istanze sono rappresentate dai nodi del grafo (rettangoli) con appropriate etichette.

un'informazione data *c* (Giunti 2010, 242). Intuitivamente, ciò significa che l'informazione *i* è tanto più rilevante per *c*, quanto più il nodo o percorso che rappresenta *i* è più vicino al nodo o percorso che rappresenta *c*.

Avendo ridotto la rilevanza a vicinanza fra le porzioni significanti (nodi o percorsi) di una rete, potremo poi sviluppare applicazioni che reperiscano automaticamente le informazioni più rilevanti per un'informazione data *c*. Tali applicazioni, a seconda di come interpreteremo l'informazione di partenza *c*, risulteranno essenzialmente di due grandi tipi. Se l'informazione iniziale *c* sarà intesa come la chiave per una richiesta, otterremo un motore di ricerca semantico. Se invece *c* sarà intesa come il punto di partenza di un viaggio nella rete, avremo un browser semantico.

Per avere un'idea di quali possano essere gli scenari futuri, assolutamente rivoluzionari, aperti dalla possibilità di effettuare vere e proprie ricerche semantiche, è opportuno confrontare questo tipo di ricerche, basate sulla rilevanza dell'informazione, con quelle usuali, che possiamo effettuare con un qualsiasi motore di ricerca tradizionale.

Come ben noto, per effettuare una qualsiasi ricerca con un motore tradizionale è necessario avere almeno un'idea di che cosa stiamo cercando. In altri termini, abbiamo bisogno almeno di una descrizione parziale (che può anche ridursi a una sola parola) dell'informazione cercata. Questa descrizione costituirà la chiave che forniremo al motore di ricerca e, tanto più dettagliata e precisa sarà la descrizione, tanto più sarà probabile che i risultati proposti siano ragionevolmente pochi e rilevanti. Inversamente, tanto più generica e vaga la descrizione, tanto più probabile che i risultati siano molti e irrilevanti.

Ciò significa che una ricerca tradizionale funziona tanto meglio quanto più conosciamo già ciò che cerchiamo; ma allora, essa è inutile in ambedue le situazioni limite - conoscenza massima o ignoranza massima dell'oggetto cercato. Questo problema, connaturato a ogni tipo di ricerca tradizionale, è ben conosciuto fin dall'antichità. Platone, nel *Menone*, ne dà una formulazione estremamente chiara, nella forma di un paradosso:

- MENONE E in quale maniera ricercherai, o Socrate, questa che tu non sai affatto che cosa sia? E quale delle cose che non conosci ti proporrai di indagare? O, se anche tu ti dovessi imbattere proprio in essa, come farai a sapere che è quella, dal momento che non la conoscevi?
- SOCRATE Capisco che cosa intendi dire, o Menone. Guarda che argomento eristico adduci: che non è possibile per l'uomo ricercare né ciò che sa né ciò che non sa! Infatti, né potrebbe cercare ciò che sa, perché lo sa già, e intorno a ciò non occorre ricercare, né ciò che non sa, perché, in tal caso, non sa che cosa ricercare. (Reale 2000, 80d-e)

In una ricerca semantica, l'informazione iniziale (chiave) che dobbiamo fornire al motore di ricerca non consiste in una descrizione, più o meno

dettagliata, di ciò che cerchiamo. Essa, al contrario, è una descrizione di ciò che noi già conosciamo su un argomento di nostro interesse. In altre parole, essa è una descrizione, più o meno dettagliata, del nostro stato conoscitivo. Ciò che noi, come soggetti individuali, sappiamo è in generale soltanto una piccola porzione della rete (un nodo e pochi nodi limitrofi ad esso collegati) di cui il motore di ricerca dispone. Una volta che forniamo al motore di ricerca il nostro stato conoscitivo, esso non farà altro che immergere questa piccola porzione nel contesto più ampio di tutta la rete, trovando le connessioni mancanti e quindi indicandoci le informazioni più rilevanti.

In altri termini, invece di dire al motore di ricerca 'trova questo e quello' potremo dire: 'su questo argomento io so questo e quello. Che cos'altro è rilevante e dovrei sapere?' Un motore di ricerca semantico sarà in grado di indicarci proprio ciò che ignoriamo assolutamente, ma esiste nella rete, ed è rilevante per ciò che già sappiamo.

3.2 Connessione e integrazione con altre basi di conoscenza: Linked Open Data e Semantic Web

Abbiamo visto come una base di conoscenze, espressa in forma ontologica, renda possibile lo sviluppo di applicazioni semantiche, che segneranno un vero e proprio salto di qualità nella ricerca e nel reperimento di informazioni. Naturalmente, il discorso fatto fin qui rispetto al caso specifico di BIA è generalizzabile ad altre basi di conoscenze, relative ad altri saperi specifici.

In generale, possiamo affermare che un'applicazione semantica sarà tanto più utile e potente quanto più la rete di conoscenze su cui opera sarà estesa e interconnessa e, soprattutto, conterrà informazioni di elevata qualità sul maggior numero possibile di domini specifici. In questa prospettiva, diventa quindi fondamentale creare le condizioni perché basi di conoscenze diverse, e di alta qualità, possano connettersi e integrarsi in modo naturale.

Tornando di nuovo, per concretezza, al caso di BIA, chiediamoci in che modo è possibile ipotizzare una connessione e integrazione di BIA con altre basi di conoscenze. A questo proposito, ricordiamo che l'ontologia di BIA è stata implementata in OWL (W3C 2004), un linguaggio ontologico che fa parte delle tecnologie del Semantic Web (Berners-Lee, Hendler, Lassila 2001; Grigoris, van Harmelen 2008). Ciò implica che, quando il popolamento dell'ontologia di BIA sarà completato, tutti i suoi dati saranno nella forma di triple RDF e quindi potranno essere resi disponibili sul web come Linked Open Data (Berners-Lee 2006, 2009).

Più in dettaglio, ciò significa anche che BIA potrà interconnettersi con qualsiasi altra base di conoscenze che sia disponibile sul web sotto forma

di Linked Open Data. Concretamente, affinché la connessione fra due basi di conoscenze si verifichi, è sufficiente che sia formalmente esplicitata almeno una relazione che colleghi le due basi. Questo processo è esattamente lo stesso di quello che ha permesso di connettere i tre schemi ontologici di BIA (par. 2.4) e quindi produrre una base di conoscenze unitaria nella forma di un grafo RDF interconnesso.

È importante sottolineare che l'interconnessione di basi di conoscenze diverse è un processo del tutto naturale e, per così dire, 'incruento', perché non mette in alcun modo in discussione la struttura interna e la specifica organizzazione concettuale delle diverse basi. Essa, al contrario, valorizza le specificità di ciascuna base, in quanto le trasforma in risorse universalmente disponibili e interoperabili.

Come accennato sopra, l'interconnessione di basi di conoscenze diverse permetterà poi di sviluppare applicazioni semantiche che potranno operare indifferentemente su tutta la rete o soltanto su porzioni specifiche di essa. Solo quando tali applicazioni saranno sufficientemente sviluppate e diffuse potremo parlare di una vera e propria integrazione fra le diverse basi di conoscenze. Più la rete sarà integrata, più gli strumenti di ricerca semantica potranno suggerire prospettive e collegamenti interessanti e inaspettati, sia agli specialisti, sia ai novizi di una data disciplina.

Bibliografia

- Berners-Lee, Tim (2006). «Linked Data» [online]. URL <http://www.w3.org/DesignIssues/LinkedData.html> (2015-07-10).
- Berners-Lee, Tim (2009). «The Next Web of Open, Linked Data» [online]. *TED2009 Conference*. URL http://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=it (2015-07-10).
- Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001). «The Semantic Web» [online]. *Scientific American*, May 17. URL <https://www.scientificamerican.com/author/tim-berners-lee-james-hendler-and-ora-lassila/> (2015-07-10).
- Burnet, John (1903). *Plato: Platonis Opera*. Edited by John Burnet. Oxford: Oxford University Press.
- Corcho, Oscar; Gómez-Pérez, Asuncion (2002). «Ontology Languages for the Semantic Web». *IEEE Intelligent Systems*, 17(1), 54-60.
- Gangemi, Aldo; Presutti, Valentina (2009). «Ontology Design Patterns». Staab, Steffen; Studer, Rudi (eds.), *Handbook of Ontologies*. 2a ed. Berlin: Springer, 221-43.
- Giunti, Marco (2010). «Grafici pesati e relazioni n-arie. Un approccio generale all'organizzazione automatica di dati secondo rapporti di rilevanza». Gola, Elisabetta; Storari, Pietro (a cura di), *Forme e formalizzazioni*. Cagliari: CUEC Editrice, 229-45.

- Giunti Marco; Sergioli, Giuseppe; Vivianet, Giuliano (2014). «L'ontologia BIA-Net. Una base per la ricerca di informazioni secondo rapporti di rilevanza nella Bibliotheca Iuris Antiqui» [online]. Ciotti, Fabio (a cura di), *Digital Humanities. Progetti italiani ed esperienze di convergenza multidisciplinare = Atti del convegno annuale AIUCD* (Firenze, 13-14 dicembre 2012). Roma: Sapienza Università Editrice, 263-78. URL http://digilab2.let.uniroma1.it/ojs/index.php/Quaderni_DigiLab/issue/view/12 (2015-07-07).
- Grigoris, Antoniou; van Harmelen, Frank (2008). *A Semantic Web Primer*. 2a ed. Cambridge (MA): MIT Press.
- Palazzolo, Nicola (1995). «Un Thesaurus per la ricerca sui diritti dell'Antichità. Esperienze e problemi». *Informatica e diritto*, 2, 267-78.
- Palazzolo, Nicola (2002). *Bibliotheca Iuris Antiqui. Sistema informativo integrato sui diritti dell'Antichità*. Include CD-ROM. Catania: Editrice Torre.
- Reale, Giovanni (2000). *Platone: Menone*. Edizione italiana a cura di Giovanni Reale, basata sull'edizione a cura di John Burnet (1903). Milano: Bompiani.
- W3C (2004a). *OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004* [online]. URL <http://www.w3.org/TR/owl-features/> (2015-07-07).
- W3C (2004b). *RDF Primer. W3C Recommendation 10 February 2004* [online]. URL <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/> (2015-07-07).
- W3C (2006). *Defining N-ary Relations on the Semantic Web. W3C Working Group Note 12 April 2006* [online]. URL <http://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/> (2015-07-07).

WordNets per lingue classiche

Riccardo del Gratta

(Istituto di Linguistica Computazionale "A. Zampolli", CNR-ILC, Pisa, Italia)

Abstract The WordNet for Ancient Greek (AGWN) is presented and illustrated both as a stand-alone semantic net and as a participant in a more complex net of wordnets for historical and modern languages. Two applications of AGWN carried out within the *Memorata Poetis* project are described: in the first application, the AGWN is used to investigate multilingual synonyms; in the second one, AGWN is used to extract Greek synonyms in order to classify epigrams in terms of similar content.

Sommario 1 Introduzione. – 2 WordNet e *WordNets*. – 3 Ancient Greek WordNet (AGWN). – 4 Applicazione di AGWN a testi poetici greci annotati con temi/motivi. – 5 Applicazione di AGWN a testi epigrafici greci. – 6 Conclusioni.

Keywords WordNet. Classical Languages. Ancient Greek WordNet. Perseus. NLP.

1 Introduzione

Il presente contributo descrive brevemente il processo di creazione di una rete di *WordNets* per le lingue classiche e ne illustra due diverse destinazioni d'uso, entrambe incentrate sul progetto *Memorata Poetis*.¹

Nella prima destinazione d'uso, partendo da testi poetici greci manualmente annotati a livello di temi e motivi si espandono le correlazioni tra parole nel testo e temi/motivi con cui i testi sono stati annotati aggiungendo sinonimi e relazioni semantiche sia alle prime che ai secondi. La seconda destinazione applica la rete semantica a epigrammi greci. Si cerca di raggruppare epigrammi che contengono termini che siano tra loro sinonimi o in relazione ipo/iperonimica. Entrambe le applicazioni sfruttano il modello concettuale delle reti semantiche: queste forniscono termini tra loro correlati a un insieme di testi da analizzare validando o meno le annotazioni o i raggruppamenti effettuati manualmente.

1 *Memoria Poetica e Poesia della Memoria*, progetto PRIN 2010/2011.

2 WordNet e WordNets

WordNet (Fellbaum 1998) è una risorsa lessico-semantiche il cui scopo principale è quello di formalizzare l'apparato teorico per creare e gestire una rete di concetti espressi da parole.

Per gli scopi del presente contributo, è sufficiente descrivere solo i blocchi principali del modello, rimandando ancora a Fellbaum (1998) per una completa ed esaustiva descrizione sia del modello completo che delle molteplici applicazioni di WordNet in diversi campi della Linguistica Computazionale.

Una tipica serializzazione del modello concettuale di WordNet è la trasformazione degli oggetti del modello in tabelle di un database relazionale. Le quattro tabelle principali e le relazioni che le connettono sono schematicamente mostrate nella fig. 1.

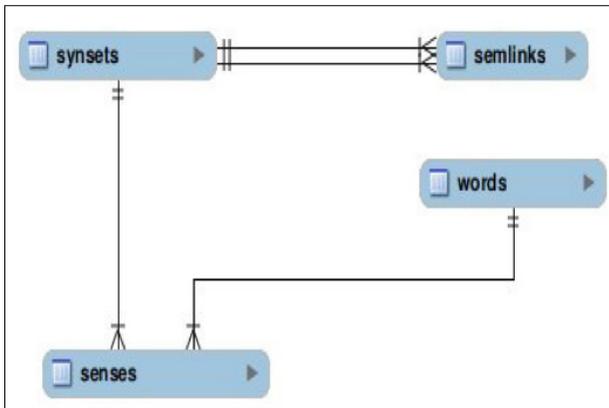


Figura 1. Le tabelle principali del modello WordNet (serializzato in un database)

La tabella 'words' contiene i singoli lemmi, come ad esempio *arm*, *limb*, *build up*, ecc. mentre i diversi sensi che i singoli lemmi possono assumere sono elencati nella tabella 'senses'. Questa tabella infatti contiene, ad esempio, il senso di *arm* sia come 'arto umano' (human limb) che come 'prepararsi per un'azione militare'. Per chiarezza, occorre dire che la tabella 'senses' non contiene le glosse dei sensi, i.e. 'arto umano' e 'prepararsi per un'azione militare', ma solo gli identificativi di queste glosse che, per contro, sono gestite dalla tabella dei 'synsets'.

La tabella 'synsets' è il cuore del modello. Oltre ad aggiungere le glosse ai sensi dei lemmi, essa svolge un doppio ruolo fondamentale: identifica univocamente i sensi aggiungendovi la parte del discorso (PoS) e raggruppa i sensi (corredati di PoS) in insiemi di sinonimi (il nome synset deriva dalla contrazione di *synonym sets*). Il senso del lemma *arm*, come verbo, inteso come 'prepararsi per un'azione militare' è sinonimo di *build up*, *fortify* e *gird*.

Infine la tabella 'semlinks' contiene le relazioni semantiche tra synsets: *arm* come arto umano è un iponimo di *limb*, inteso come una delle quattro appendici del corpo umano. La fig. 2 riporta graficamente gli esempi sopra descritti.

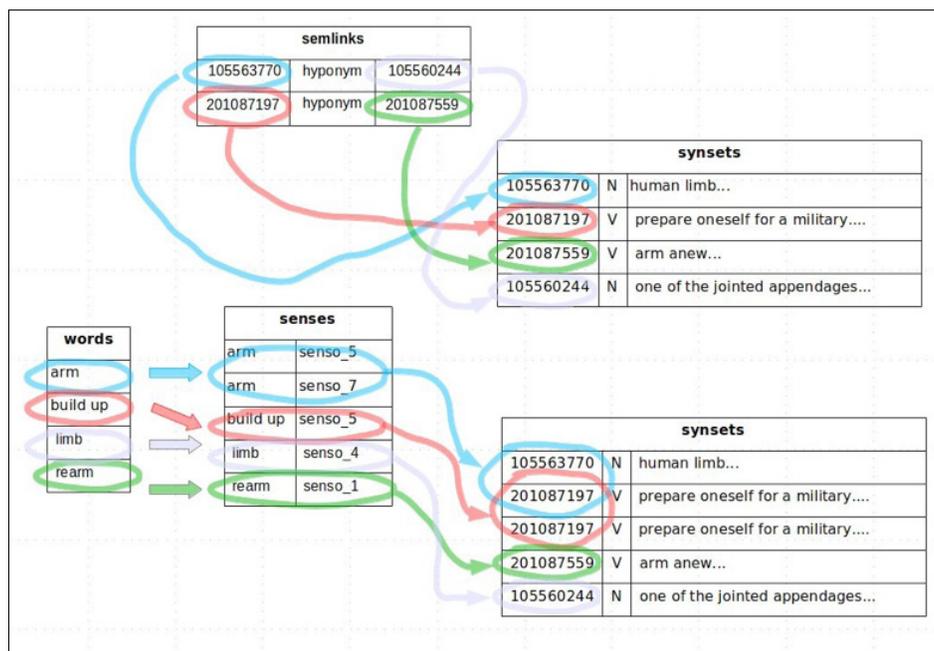


Figura 2. Parole, sensi, synsets e relazioni semantiche

Sebbene WordNet sia stato pensato espressamente per concettualizzare la lingua inglese (in letteratura viene citato come Princeton WordNet), il modello è generico e può essere applicato a lingue diverse, sia europee (Vossen 1998) che semitiche (Rodriguez et al. 2008). Seguendo il lavoro in Minozzi (2009), Bizzoni et al. (2014) hanno codificato la creazione di un database lessico-semantic per il greco antico basato sul modello WordNet.

3 Ancient Greek WordNet (AGWN)

Come mostrato in Bizzoni et al. (2014), la costruzione iniziale della WordNet per il greco antico si basa su dizionari bilingui digitalizzati greco-inglese forniti dal progetto Perseus (<http://www.perseus.tufts.edu>). La metodologia è quella di estrarre coppie di parole greco-inglese e confrontare la parola inglese con la lista di lemmi contenuti in Princeton WordNet. Se la

parola inglese è trovata, i suoi synsets vengono assegnati alla corrispondente parola greca.

Essendo il processo di creazione un processo di bootstrapping, ovvero di estrazione non controllata di dati, il risultato risulta alquanto impreciso: la polisemia indotta dall'inglese porta, ad esempio, la parola greca ἔχω a essere mappata su oltre 170 parole inglesi, tra cui *cut*, *make*, *break*; mentre se la parola inglese della coppia non è presente su Princeton WordNet, la parola greca viene esclusa, fatto, questo, che ha come conseguenza di limitare la copertura della WordNet per il greco antico al 30% del lessico totale.

Al momento attuale, la WordNet del greco antico è mappata su latino, italiano e inglese, in modo da avere una rete di wordnet interconnesse; anche questa mappatura, come quelle in Vossen (1998), Rodriguez et al. (2008), sfrutta l'inglese come lingua pivot, ovvero come lingua scelta per connettere tra di loro le altre lingue, per cui, di conseguenza, il greco e il latino o l'italiano sono interconnessi attraverso l'equivalente senso inglese.

4 Applicazione di AGWN a testi poetici greci annotati con temi/motivi

Lo scopo di questa applicazione è verificare da un punto di vista semantico se l'annotazione di un dato testo poetico greco con un particolare tema/motivo (in latino) sia suffragata da un'evidenza semantica.

Nell'applicazione si associano statisticamente una o più parole greche del testo con il tema/motivo con cui il testo è stato manualmente annotato. Questa lista di associazioni (le quali sono coppie bilingue greco-latine) viene controllata su AGWN per capire se la parola greca e quella latina sono mappate sullo stesso synset inglese. In altre parole, se la parola greca e quella latina sono sinonimi multilingui.² Il fatto che le parole greche e latine siano mappate sullo stesso synset inglese significa che sono entrambe traducenti dello stesso concetto ed in questo caso l'annotazione è molto significativa. La metodologia è ampiamente descritta in Boschetti (2014), a cui si rimanda anche per ulteriori casi interessanti. Di seguito è riportato un caso prototipico di sinonimia multilingue.

Da un testo si estrae la coppia termine greco/tema (in latino) {γραιῶς, anus} che risulta mappata sullo stesso synset inglese che esprime il concetto di "donna vecchia".³ In questo caso l'annotazione del testo con il

2 Spesso la parola greca e quella latina non sono sinonimi multilingui, ma c'è comunque evidenza di una connessione a livello di iperonimi: la parola greca è sinonima multilingue di una parola latina coiponima di quella sotto indagine, oppure gli iperonimi della coppia sono sinonimi multilingui.

3 In Princeton WordNet 3.0 il synset è identificato dal numero univoco 110377021.

tema/motivo *Anus* risulta potenziata dal fatto che la parola greca staticamente più correlata al tema/motivo (γραῦς) esprime esattamente lo stesso concetto espresso dalla parola latina *anus*.

In aggiunta, la lista di sinonimi greci e latini che si estraggono dal synset comune permettono una associazione tra una qualsiasi parola greca sinonima di γραῦς e una qualsiasi parola latina sinonima di *anus*:

{γραίδιον, γραῖα, γραῦς, τηθία} → {*anicla, anicula, anicilla, anucella, anus*}

5 Applicazione di AGWN a testi epigrafici greci

La seconda applicazione d'uso è il raggruppamento di epigrammi greci. Attraverso AGWN, l'applicazione permette di raggruppare epigrammi che contengono termini tra loro sinonimi o in relazione (ipo)iperonimica tra di loro. Ad esempio i tre epigrammi, AG 9,492,493,494, sotto

- Κεῖτο δ' ὁμοῦ σάκος, ἔγχος, <ἄορ>, θώρηξ, κόρυς, ἵππος.
- Ἀσπίς, τόξα, βέλεμνα, κόρυς, ξίφος, ἄλκιμον ἔγχος.
- Ἴος, τόξα, σάκος, κυνέη, δόρυ, φάσγανα, θώρηξ.

trattano argomenti simili e possono essere raggruppati come ricorrenza lessico-semantica nella versificazione epigrafica che è un aspetto centrale del progetto Memorata Poetis.

Infatti AGWN rileva automaticamente la sinonimia tra σάκος e ἀσπίς, nel significato di scudo (shield); ξίφος e φάσγανα, intesi come spada (blade, sword) e ξίφος e ἄορ ancora come spada (blade, sword).

6 Conclusioni

In conclusione, si è cercato di illustrare come la WordNet per il greco antico, intesa sia come rete semantica a sé stante, sia come rete interconnessa a reti semantiche per le lingue classiche e moderne, sia in grado di aiutare l'annotazione e la classificazione di testi in greco antico. L'annotazione di testi poetici a livello di temi e motivi risulta potenziata quando le parole greche contenute nel testo e maggiormente correlate alla parola latina esprime il tema/motivo sono semanticamente connesse; gli epigrammi risultano più facilmente classificabili per temi trattati grazie all'estrazione di sinonimi in essi contenuti. Nel primo caso la WordNet per il greco antico coopera con le altre reti semantiche, specificatamente quella latina e quella inglese, mentre nel secondo caso la rete è usata come risorsa lessico-semantica a sé stante.

Bibliografia

- Bizzoni, Yuri et al. (2014). «The Making of Ancient Greek WordNet». *Proceedings of the 9th Conference on Language Resources and Evaluation* (Reykjavik, 26th-31st May 2014). Paris: ELRA, 1140-7.
- Boschetti, Federico; Del Gratta, Riccardo (2014). «Computer Assisted Annotation of Themes and Motifs in Ancient Greek Epigrams. First Steps». Basili Roberto; Lenci, Alessandro; Magnini, Bernardo (eds.), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014* (Pisa, 9-10 December 2014). Pisa: Pisa University Press, 83-6.
- Fellbaum, Christiane (1998). *WordNet. An Electronic Lexical Database*. Cambridge (MA): The MIT Press.
- Minozzi, Stefano (2009). «The Latin WordNet Project». Anreiter, Peter; Kienpointner, Manfred (eds.), *Latin Linguistics Today = Akten des 15. Internationalen Kolloquiums zur Lateinischen Linguistik* (Innsbruck, 4-9 April 2009). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck Bereich Sprachwissenschaft, 137: 707-16.
- Rodríguez, Horacio et al. (2008). «Arabic WordNet. Current State and Future Extensions». *Proceedings of the Fourth International Global WordNet Conference. GWC 2008* (Szeged, Hungary, 22nd-25th January 2008). Amsterdam: IOS Press, 387-406.
- Vossen, Piek (1998). *EuroWordNet. A Multilingual Database with Lexical Semantic Networks*. Norwell (MA): Kluwer Academic Publishers.

Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'Information Retrieval

Stefano Minozzi

(Università degli Studi di Verona, Italia)

Abstract This paper describes the creation of Latin WordNet, addressing the methods of construction, the data structure, possible applications and further developments. More notably, a possible expansion of Latin WordNet is the construction of an additional data structure in order to improve the chances of the network to represent the semantic shift taking diachrony into account. The original Princeton WordNet, on whose ground Latin WordNet is developed, does not take into account the semantic shift of words in time, considering the meanings identified by synsets as if existing in a model of vocabulary where a language has not an history: the purpose of Princeton WordNet was obviously the representation of in-use contemporary English, but this model is weak when we come to describe the semantic structures of a finished language such as Latin. In the second part, this paper shows the connections between semantic networks and information retrieval moving from the problems pertaining to the automatic disambiguation of contexts and the exploitation of ontologies in the process of sense retrieval on semantically annotated corpora.

Sommario 1 La costruzione della base di conoscenza lessicale LatinWordNet. – 2 Ipotesi di utilizzo di una rete basata su WordNet per l'*Information Retrieval*.

Keywords WordNet. Ontologies. Lexicography. Dictionaries. Semantic network. Information retrieval.

1 La costruzione della base di conoscenza lessicale LatinWordNet

Il progetto di costruzione di una base di conoscenza lessicale per la lingua latina nacque, alla fine del 2004, con l'intento di fornire un modello di rappresentazione elettronica della conoscenza semantica utile alla applicazione di tecniche di *Natural Language Processing* per quella lingua.

Uno degli esempi più completi di rete semantica, all'epoca, era costituito da WordNet (Miller et al. 1990; Fellbaum 1998), un sistema disponibile pubblicamente, che contiene *frame* specificamente orientati alla rappresentazione delle parole: a partire dal riconoscimento della natura del tutto accidentale dell'ordinamento dei dizionari attraverso spelling, nel modello

di WordNet le parole sono organizzate per blocchi di significato, denominati 'synsets', che raccolgono tutti i lemmi che lessicalizzano lo stesso concetto; i synsets sono collegati tra loro per mezzo di relazioni che includono, assieme alla sinonimia, anche l'iponimia, la meronimia e l'antinomia. L'ipo-iperonimia mette in relazione significati subordinati e superordinati fornendo così una struttura gerarchica di concetti. La relazione meronimica induce una gerarchia delle parti sull'insieme dei significati. In questo modo il livello lessicale è chiaramente separato da quello concettuale e questa distinzione è rappresentata dal medium semantico-concettuale e dalla relazione semantica che uniscono rispettivamente synsets e parole. Le relazioni presenti tra i verbi permettono di mettere in luce relazioni di implicazione (ingl. 'entailment') e di troponimia. Due verbi sono correlati dall'implicazione nel momento in cui il primo verbo implichi il secondo: per esempio la coppia comprare-pagare. La troponimia è la relazione presente nel momento in cui due attività collegate da implicazione avvengono allo stesso tempo: un esempio è la coppia zoppicare-camminare.

Il progetto di costruzione di una rete semantica per il latino ebbe come basi di partenza due modelli: WordNet e MultiWordNet (Bentivogli, Girardi, Pianta 2002), progetto sviluppato dall'allora Istituto Trentino di Cultura (oggi Fondazione Bruno Kessler) inteso a realizzare una rete semantica multilingue. Il modello adottato dal progetto MultiWordNet (MWN) consisteva nel costruire le reti semantiche specifiche per un linguaggio mantenendo il più possibile le relazioni semantiche disponibili nella WordNet di Princeton (PWN). Ciò è ottenibile costruendo i nuovi synsets in corrispondenza dei synsets della PWN, ogni volta che ciò sia fattibile, e importando le relazioni semantiche dai corrispondenti synsets inglesi; in questo modo si ipotizza che, se esistono due synsets nella PWN e una relazione che li collega, la stessa relazione leghi i corrispondenti synsets in una lingua diversa. Secondo Vossen (1996), il modello di MWN (o 'modello a espansione', *expand model*) garantisce un elevato grado di compatibilità tra differenti *wordnet*. Per constatare questo fatto basta considerare che la costruzione di qualsiasi rete semantica necessariamente implica un gran numero di decisioni soggettive (e discutibili). Così se due reti semantiche sono costruite indipendentemente per due diverse lingue, mostreranno differenze che dipendono solo parzialmente dalle differenze tra le due lingue: alcune non banali discrepanze strutturali dipenderanno, infatti, da scelte soggettive o da criteri di costruzione differenti. Il modello di MWN minimizza queste differenze aderendo strettamente ai modelli di costruzione di PWN.

Il modello MWN presentava anche degli inconvenienti potenziali: il rischio più serio era quello di forzare un'eccessiva dipendenza sulla struttura lessicale e concettuale di uno dei linguaggi coinvolti (Vossen 1996). Questo rischio però poteva essere scongiurato permettendo alla nuova rete semantica di divergere, quando necessario, dalla struttura di PWN.

Un altro importante vantaggio del modello MWN era la possibilità di utilizzare procedure automatiche per velocizzare la costruzione dei synsets corrispondenti e per l'individuazione delle divergenze tra PWN e la rete semantica che si stava costruendo. In tutte queste procedure la stessa PWN poteva essere usata utilmente come risorsa.

La costruzione di LWN (LatinWordNet) si è pertanto basata in un primo tempo su una procedura automatica per l'assegnazione. Seguendo il modello MWN, il nostro obiettivo era quello di costruire, ogniqualevolta fosse possibile, un synset latino sinonimo (semanticamente corrispondente) di un synset di PWN. Se ciò non era possibile, veniva individuata un'idiosincrasia Inglese-a-Latino o Latino-a-Inglese.¹

I synsets sinonimi latini sono stati costruiti seguendo tre differenti strategie:

- La prima strategia era basata sui traducenti dall'inglese al latino. Per ciascun synset di PWN *S*, cerchiamo un gruppo di traducenti che siano i sinonimi delle parole inglesi di *S*. Se non è possibile costruire alcun synset sinonimo latino di *S*, si è trovata un'idiosincrasia lessicale inglese-a-latino.
- La seconda strategia era basata sui gruppi di traducenti Latino-a-Inglese. Per ciascun senso σ di una parola latina *L*, si cerca un synset di PWN che includa almeno un traduceente inglese di *L* e si costituisce un legame tra *L* e *S*. Quando la procedura è stata applicata a tutti i significati della parola latina, possiamo costruire la classe di equivalenza di tutti i gruppi di parole latine che sono state collegate con lo stesso synset di PWN. Ciascun gruppo nella classe di equivalenza è il synset latino sinonimo con alcuni synsets di PWN. Se per un gruppo di sinonimi latini non c'è alcun synset sinonimo in PWN, si è trovata un'idiosincrasia lessicale Latino-a-Inglese.
- La terza strategia sfruttava la natura multilingue di MWN e i due dizionari di macchina latino-inglese e latino-italiano: attraverso di essa le parole latine che risultano avere come traduzione parole inglesi e parole italiane che sono contrassegnate dallo stesso identificativo di *synset* vengono attribuite al medesimo *synset* della rete LatinWordNet con lo stesso identificativo. Esse infatti presentano con certezza la lessicalizzazione latina del concetto espresso dai traducenti nelle due lingue moderne.²

¹ Una trattazione estesa relativa al problema dei *lexical gap* si può trovare in Bentivogli et al. 2000.

² In altre parole una parola, data una parola latina, sarà attribuita a quei *synset* che costituiscono l'intersezione dei gruppi di *synset* individuati dai rispettivi gruppi di traducenti: vale a dire quei *synset* ai quali rimandano sia i traducenti italiani, sia i traducenti latini.

Il miglior allineamento tra la WordNet di Princeton e quella latina è stato ottenuto utilizzando entrambe le strategie per cercare di validare i risultati incrociandoli.

Trovare collegamenti tra i significati delle parole latine e i synsets di PWN è un processo complesso e lungo, anche se è sempre molto più rapido rispetto alla costruzione da zero dei synsets latini, della loro organizzazione in una rete semantica e del metterli in corrispondenza con i synsets di PWN. Per ciascun significato latino, il lessicografo dovrebbe cercare i gruppi di traducenti equivalenti in un dizionario bilingue, trovare tutti i synsets che contengono questi traducenti equivalenti, valutare con attenzione il significato di questi synsets (sinonimi, glosse, relazioni semantiche) e, infine, decidere quale *synset* di PWN, se esiste, è sinonimo del significato latino della parola. Per alcuni significati di parola il lessicografo potrebbe dover valutare decine di synsets di PWN.

Per aiutare il lessicografo nel suo lavoro era stata realizzata una procedura che sceglieva, per ciascun significato di una parola latina, i synsets di PWN dal significato compatibile. Nella maggior parte dei casi la procedura trovava una rosa ristretta di candidati, aiutando il lessicografo a focalizzare quale fosse il *synset* di PWN più per l'assegnazione.

La procedura-assegnazione prendeva come input uno dei sensi della sezione Latino-a-Inglese del dizionario di macchina e forniva in output un gruppo di candidati, ciascuno dei quali era descritto da un *punteggio di certezza* e da un *synset* di PWN, dove il punteggio di certezza (PC) misurava il grado di certezza nel legame tra il significato della parola latina e il *synset* di PWN. Solo i candidati con un PC più alto di una certa soglia venivano proposti al lessicografo. Scegliere il livello di soglia è stata una questione di bilanciare precisione e richiamo. Maggiore era la soglia, minore era la probabilità che candidati erronei fossero proposti (alta precisione), ma era anche maggiore la possibilità che la scelta più idonea non fosse inclusa nel gruppo dei candidati (basso richiamo).

Per un determinato significato di parola listato nel dizionario latino-inglese, la procedura-assegnazione considerava il gruppo di parole inglesi che venivano proposte come traducenti equivalenti per quel significato e trovava tutti i synsets contenenti almeno un traducevole equivalente. Questi *synsets* costituivano il gruppo di candidati (GCand) che doveva essere collegato con il significato di parola latina dell'input. Possiamo riassumere il primo passo dell'algoritmo dicendo che esso calcolava i GCand del significato di una determinata parola latina. Il resto dell'algoritmo consisteva nell'ordinare i GCand calcolando il PC di ciascuno dei synsets.

L'ordinamento dei GCand era basato su una serie di regole per stabilire i legami: ogni regola, se applicata con successo a un candidato, alzava il suo PC. Si deve notare che il PC parziale, contribuito da ciascuna regola, variava a seconda di fattori specifici alla regola. Accanto al dizionario di macchina, venivano utilizzate dalle regole anche altre risorse, come la se-

zione italiana di MultiWordNet e un dizionario italiano-latino, un dizionario dei sinonimi latini e la stessa PWN.

Le strategie individuate per la costruzione dei legami erano quattro:

- probabilità generica: la regola di probabilità generica si basa sulla supposizione che solo un elemento nel GCand è il corretto candidato per legare il senso di una parola latina. Di conseguenza si può supporre che maggiore è la cardinalità del GCand, minore è la probabilità che ciascun candidato sia quello esatto. La cardinalità del GCand dipende dal grado di ambiguità delle parole che sono proposte come traducenti equivalenti del significato della parola di input. Se c'è un solo synset nel GCand, ciò significa che tutti i traducenti equivalenti della parola di input sono monosemici: è quindi altamente probabile che l'unico synset nel GCand sia sinonimo del significato della parola di input.³
- traduzione incrociata: questa regola si basa sulla supposizione che se colleghiamo un significato di parola al corretto synset attraverso un traducente equivalente, è probabile che almeno alcuni dei sinonimi del traducente, presenti in PWN, abbiano la parola di input come traducente equivalente inglese-latino. Si prenda ad esempio il latino 'punctum': quando riferito a insetti, si traduce come 'sting'. 'Sting', però, appartiene a 4 synsets di PWN: 'sting', 'stinging'; 'pang', 'sting'; 'sting', 'bite', 'insect bite'; 'bunco', 'bunco game', 'sting'. Solo il terzo synset è sinonimo della parola latina. Se guardiamo ai sinonimi di 'sting' nel terzo synset possiamo trovare che la sezione inglese-latino dà 'punctum' come traduzione di 'bite'. Riassumendo, la regola della traduzione incrociata considera i sinonimi presenti in PWN di un traducente che crea il collegamento e calcola un PC parziale che è proporzionale al numero di sinonimi che hanno la parola italiana come traducente dall'inglese al latino.
- corrispondenza della glossa: un gruppo di regole di collegamento sfrutta le informazioni contenute nella glossa inglese che introduce la maggior parte del dizionario di macchina inglese-latino. La glossa può contenere un campo semantico specifico, un sinonimo, un iperonimo, o una specificazione di contesto d'uso. Queste informazioni possono essere utilizzate in vario modo.

L'informazione relativa al campo semantico è sfruttata grazie a una risorsa sviluppata parallelamente a MWN, cioè la marcatura di tutti i synsets di PWN con una etichetta relativa al campo semantico (Magnini, Cavaglià 2000). La glossa del dizionario infatti contiene una etichetta relativa al

3 Cf. il criterio monosemico usato in Atserias et al. 1997.

campo semantico e se questa etichetta corrisponde a un synset individuato come candidato, allora il candidato ottiene un maggiore PC. Le varianti nelle etichette dei campi semantici sono gestite attraverso una tabella di corrispondenze.

Quando le glosse contengono parole o frasi, si cerca un corrispondente tra di esse e le parole contenute nelle glosse di PWN. Per fare ciò, si estraggono i lemmi delle parole inglesi delle glosse, e si controlla la loro presenza nelle glosse del traduttore equivalente in PWN. La forza della corrispondenza dipende dal grado di ambiguità del traduttore. Maggiore è la polisemia, minore è il peso attribuito alla corrispondenza.

Il meccanismo ha due estensioni basate sul fatto che le glosse spesso specificano il genere della parola che stanno definendo al posto di un sinonimo. La prima estensione cerca una corrispondenza tra una parola latina e un iperonimo del suo traduttore equivalente. Il secondo meccanismo cerca una corrispondenza tra una parola latina e una parola inglese contenuta nella glossa di un iperonimo del synset candidato. Se la corrispondenza tra la parola latina e la parola inglese viene ottenuta attraverso uno dei meccanismi indiretti il PC parziale sarà più basso rispetto all'individuazione diretta.

La regola dell'intersezione di synsets sfrutta il fatto che i gruppi di traduzione possono includere più traduttori equivalenti, che sono ovviamente sinonimi. Se uno dei traduttori equivalenti è ambiguo, possiamo usare gli altri traduttori equivalenti per disambiguare. In pratica, la regola prende i differenti gruppi di candidati che sono accessibili attraverso diversi traduttori equivalenti e li interseca. I synsets che sono nell'intersezione ottengono un PC. Per esempio la parola latina 'pila' è tradotta nel suo senso metaforico come 'pillar', 'mainstay'. La parola 'pillar' appartiene a 5 synsets di PWN, mentre 'mainstay' appartiene a tre synsets. C'è però un solo synset che li contiene entrambi.

Una volta terminata l'assegnazione si è ottenuta una struttura reticolare che forniva una modellizzazione dei rapporti semantici tra le parole: il modello di stoccaggio dei dati in *MultiWordNet* rifletteva i principali elementi teorici della rete semantica multilingue. Il database era costruito sull'idea che esista un gruppo di dati comuni a tutte le lingue e altri specifici di ciascuna lingua. Nell'implementazione le relazioni semantiche di PWN sono contenute in un modulo chiamato COMMON-DB, mentre le relazioni lessicali per il latino e per l'inglese sono conservate in altri due moduli LATIN-DB e ENGLISH-DB. In altre parole l'informazione relativa a quali lemmi appartengano ai synsets si trova nei database delle lingue, mentre l'informazione relativa alle relazioni tra i *synsets*, che rimangono costanti tra le lingue, è immagazzinata nel COMMON-DB. La corrispondenza tra i synsets realizzati nelle diverse lingue si ottiene utilizzando sempre lo stesso codice identificatore: i *synsets* di lingue diverse che hanno lo stesso codice di identificazione appartengono al medesimo multisynset.

Il COMMON-DB descrive le relazioni tra i *multisynsets* di MWN. Quindi, tutte le informazioni semantiche che sono indipendenti dalla lingua possono essere aggiunte al COMMON-DB.⁴

Il modello di dati di MWN rappresenta le costanti concettuali presenti in lingue differenti. Tale modello di dati, inoltre, evidenzia anche le divergenze semantiche tra le lingue.⁵ Inoltre, anche se si mantengono le relazioni semantiche evidenziate da PWN come base del COMMON-DB, è possibile aggiungere nuove relazioni o modificare quelle esistenti. La possibilità di modificare le relazioni semantiche di PWN e di rappresentare le idiosincrasie concettuali nei linguaggi specifici è stata implementata attraverso dei moduli aggiuntivi che sovrascrivono, senza modificarli fisicamente, i dati originali di PWN. Il COMMON-DB infatti contiene tutte le relazioni semantiche originali di PWN e una risorsa chiamata COMMON-ADD-ON che ne riscrive una parte. Ciascuna lingua contiene un language-ADD-ON che specifica le relazioni semantiche che sono proprie di quella lingua.

Le peculiarità lessicali vengono codificate all'interno delle aggiunte specifiche di ciascuna lingua. Se c'è prova che la lessicalizzazione di un determinato concetto manchi in una lingua, nella sezione lessicale del database di quella lingua viene inserita un'etichetta vuota per quel nodo.⁶ Per la rappresentazione delle differenze denotative e dei gap lessicali, vengono seguite due diverse strategie: se il nodo vuoto corrisponde a una differenza denotativa, una o più relazioni vicine vengono usate per collegare il nodo ad un synset più generico o a molti *synsets* più specifici. Se il nodo vuoto corrisponde a un gap lessicale, viene riportata nella glossa del nodo vuoto una parafrasi di traduzione appropriata, preceduta dalla parola chiave TE (Translating Equivalent). Le relazioni più vicine vengono inserite nella risorsa linguistica aggiuntiva specifica della lingua in questione.

Ciascun database linguistico contiene anche un modulo con informazioni lessicografiche relative ai collegamenti tra i sensi delle parole e i synsets.

Per quel che riguarda le relazioni, tutte quelle semantiche erano state importate da PWN e sono disponibili assieme alle relazioni più vicine, cioè le nuove relazioni specifiche di ciascuna lingua che sono state aggiunte nella MWN per rappresentare le differenze denotative.

L'attuale implementazione della parte latina di MWN si basa sull'aggiunta di un modulo (non disponibile online) in grado di rendere indipendente il livello grafico/ortografico dall'individuazione dei lemmi. In pratica per ciascun lemma di dizionario è stata introdotta una grafia normalizzata,

4 In particolare le relazioni relative ai campi semantici.

5 Nella fattispecie i gap lessicali.

6 Il termine 'nodo' è usato in quanto MWN si compone di una struttura reticolare, dove i lemmi sono inseriti come nodi e le relazioni semantiche costituiscono i collegamenti tra i nodi.

associata a un numero espandibile di grafie alternative. Nei synsets della parte latina non vengono registrati direttamente i lemmi, ma dei codici identificativi: in questo modo possono essere utilizzate diverse grafie per la rappresentazione dello stesso lemma, e sono inoltre collegate all'interno del synset anche tutte le realizzazioni morfologiche della flessione dei lemmi. L'implementazione della base dati è stata effettuata attraverso un database relazionale, in modo da permettere l'interfacciamento con il sistema di IR in maniera versatile, sfruttando le possibilità di consultazione anche attraverso un ambiente distribuito.

La consistenza della base dati è di 9.378 lemmi collocati in 8.973 synsets con 143.701 archi di relazione: la copertura lessicale e i risultati dell'assegnazione automatica necessiterebbero di una ulteriore fase di valutazione e di controllo. Lo strumento è stato reso consultabile attraverso il sito della Fondazione Bruno Kessler⁷ che ha sviluppato e messo a disposizione l'interfaccia per effettuare il browsing della rete semantica latina conteo stualmente a quelle realizzate per altre lingue.

Uno dei punti deboli che si ravvisano nella trattazione di una lingua conclusa con il sistema di una rete lessicale è la mancanza di una collocazione diacronica dei significati all'interno delle stadiazioni della lingua in esame: un'interessante espansione per LWN potrebbe essere quella di un'ulteriore marcatura delle parole afferenti ai gruppi sinonimici che possa far tenere traccia dello spostamento semantico in una prospettiva temporale di storia della lingua. Tale marcatura permetterebbe di identificare l'ingresso di una parola all'interno di un synset (o la sua uscita) nel tempo. Questo lavoro di marcatura storica non può che avvenire, al momento, che attraverso la revisione delle singole parole e dei singoli nodi da parte del lessicografo.

2 Ipotesi di utilizzo di una rete basata su WordNet per l'*Information Retrieval*

Uno dei problemi delle tecniche di reperimento dell'informazione basate sulla corrispondenza di parole è determinato dal fatto che la corrispondenza tra parole e concetti non è una funzione in senso matematico. Nel caso di omografi le parole che sembrano uguali rappresentano concetti diversi, nel caso dei sinonimi, invece, due parole distinte rappresentano lo stesso concetto. Gli omografi rappresentano un ostacolo in quanto diminuiscono la precisione creando falsi positivi e i sinonimi diminuiscono i valori di richiamo in quanto si presentano come falsi negativi. Nel costru-

⁷ URL <http://multiwordnet.itc.it/online/multiwordnet.php> (2017-10-19).

ire un sistema di reperimento dell'informazione si parte dall'assunto che l'efficacia degli algoritmi di ricerca dovrebbe migliorare se il confronto non viene operato direttamente sulle parole ma sui concetti che le parole rappresentano.

Le basi di conoscenza lessicale costruite sul modello di WordNet definiscono un concetto attraverso il synset. Pertanto lo sfruttamento dell'informazione semantica contenuta in WordNet è stato investigato in vari modi, soprattutto per quel che riguarda l'interazione tra synsets e concetti nelle operazioni di ricerca testuale. Sussna (1993) e Richardson (1994) trattano la struttura creata dai puntatori relazionali di WordNet come una rete semantica e definiscono alcuni modelli di misurazione per calcolare la distanza tra synsets. La somiglianza tra una *query* e un documento viene poi calcolata dalla similarità tra il gruppo di synsets della *query* e i synsets presenti nei documenti. Quest'ultimo tipo di operazione è particolarmente esosa in termini di risorse di calcolo a causa della sua natura combinatoria: vengono infatti individuate tante coppie di synsets quanti sono gli elementi della *query* e quanti sono i documenti e devono essere valutati tutti i percorsi possibili tra ciascuna coppia.

Per contenere e ottimizzare il numero di confronti, negli esperimenti di Chakravarthy (1994) e di Chakravarthy e Haase (1995), WordNet è stata utilizzata per trovare corrispondenze quando sia le *query* sia i documenti sono brevi e strutturalmente prevedibili. Questo interessante approccio è stato applicato all'indicizzazione delle didascalie in collezioni di immagini. La rappresentazione delle *query* e delle didascalie viene realizzata automaticamente e identifica il ruolo delle parole nel testo: il confronto individua corrispondenze se le parole della *query* e quelle della didascalia sono collegate semanticamente nel *database* lessicale e rivestono lo stesso ruolo nei rispettivi testi.

Un terzo tentativo di confronto tra concetti e *synsets* è motivato dall'obiettivo di migliorare l'efficienza del sistema di ricerca, mantenendo la robustezza e l'affidabilità del modello vettoriale (Sussna 1993). Il focus di questo approccio è dato da una procedura di indicizzazione completamente automatica progettata per scegliere un solo synset per ciascuna parola nel testo. Il risultato di questa procedura di indicizzazione è un vettore nel quale alcuni dei termini rappresentano i synsets anziché le parole. Una volta creato un vettore basato sui synsets, esso viene gestito esattamente come uno basato sulle parole (vale a dire utilizzando l'analisi del coseno come parametro su cui calcolare la similarità fra richiesta e risultato restituito).

Un'alternativa di sfruttamento della rete semantica atto soltanto ad aumentare il richiamo dei risultati all'interno del sistema di ricerca può essere basato sull'utilizzo della struttura dei synsets per espansione della *query* utente. Le strategie di espansione possono essere molteplici: espansione solo attraverso i sinonimi, espansione attraverso i sinonimi e i legami di parentela nella gerarchia, espansione attraverso i sinonimi e

tutti i synsets direttamente collegati (aventi distanza 1 all'interno della catena dei collegamenti).

Gli esperimenti generalmente condotti hanno portato a rilevare come il livello di precisione delle *query* espanse sia direttamente proporzionale al numero di parole specifiche presenti nella *query* originale: l'espansione attraverso i synsets dovrebbe essere combinata in modo da operare sull'intersezione degli insiemi costituiti dai risultati delle *query* individuali. L'intersezione in genere porta a un aumento del richiamo nella maggior parte dei casi ma con una certa degradazione della precisione.

Un sondaggio più approfondito di queste ipotesi implicherebbe in ogni caso il completamento della rete semantica LatinWordNet, onde coprire con continuità tutta la catena ipo-iperonimica e la realizzazione di un vasto corpus di testi semanticamente annotati. Campo di fertile ricerca sarebbe senz'altro l'investigazione di modalità di attribuzione automatica e disambiguazione che superassero quanto già operato da Ellen M. Voorhees (1998) nei suoi ormai classici scritti: ampi orizzonti di ricerca per cui si auspica che questo convegno possa contribuire ad aprire sentieri.

Bibliografia

- Atserias, Jordi et al. (1997). «Combining Multiple Methods for the Automatic Construction of Multilingual Wordnets». Mitkov, Ruslan; Nicoulov, Nicolas (eds.), *Recent Advances in Natural Language Processing II. Selected Papers from Ranlp '97*. Amsterdam; Philadelphia: John Benjamins, 327-40.
- Bentivogli, Luisa; Pianta, Emanuele (2000). «Looking for Lexical Gaps». Heid, Ulrich (ed.), *Proceedings of the Ninth EURALEX International Congress* (Stuttgart, 8th-12th August 2000). Universität Stuttgart: Institut für Maschinelle Sprachverarbeitung, 663-9.
- Bentivogli, Luisa; Pianta, Emanuele; Pianesi, Fabio (2000). «Coping with Lexical Gaps When Building Aligned Multilingual Wordnets». Gavrilidou, Maria et al. (eds.), *2nd International Conference on Language Resources and Evaluation = Proceedings of LREC-2000* (Athens, 31st May-2nd June 2000). Paris: European Language Resources Association, 993-7.
- Bentivogli, Luisa; Girardi, Christian; Pianta, Emanuele (2002). «MultiWordNet. Developing and Aligned Multilingual Database». Central Institute of Indian Languages (ed.), *Proceedings of the First International Conference on Global WordNet* (Mysore, India, 21st-25th January 21-5). Mysore: Central Institute of Indian Languages, 293-302.
- Chakravarthy, Anil S. (1994). «Toward Semantic Retrieval of Pictures and Video». Centre de hautes études internationales d'informatique documentaire (ed.), *Intelligent Multimedia Information Retrieval Systems*

- and Management = *RIAO '94 Conference Proceedings with Presentation of Prototypes and Operational Systems*. Paris: CID-CASIS, 1: 676-86.
- Chakravarthy, Anil S.; Haase, Ken B. (1995). «Net.Serf. Using Semantic Knowledge to Find Internet Information Archives». *SIGIR '95 = Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Seattle, 9th-13th July 1995). New York: ACM Press, 4-11.
- Fellbaum, Christiane (1998). *WordNet. An Electronic Lexical Database (Language, Speech, and Communication)*. Cambridge: MIT Press.
- Gavriliidou, Maria et al. (eds.) (2000). *2nd International Conference on Language Resources and Evaluation = Proceedings of LREC-2000* (Athens, 31st May-2nd June 2000). Paris: European Language Resources Association.
- Magnini, Bernardo; Cavaglià, Gabriele (2000). «Integrating Subject Field Codes into Wordnet». Gavriliidou et al. 2000, 1413-18.
- Miller, George A. et al. (1990). «Introduction to WordNet. An on-Line Lexical Database». *International Journal of Lexicography*, 3(4), 235-44.
- Richardson, Ray (1994). *A Semantic-Based Approach to Information Processing* [PhD Dissertation]. Dublin: Dublin City University.
- Sussna, Michael (1993). «Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network». *CIKM '93 = Proceedings of the Second International Conference on Information and Knowledge Management*. New York: ACM Press, 67-74.
- Voorhees, Ellen M. (1993). «On Expanding Query Vectors with Lexically Related Words». Harmna, Donna (ed.), *Proceedings of the First Text REtrieval Conference (TREC-1)* (Gaithersburg, Maryland, November 1-3 1995). Gaithersburg: National Institute of Standards and Technology, 223-32.
- Voorhees, Ellen M. (1998). «Using WordNet for Text Retrieval». Fellbaum 1998, 285-303.
- Vossen, Piek (1996). «Right or Wrong. Combining Lexical Resources in the Eurowordnet Project». Gellerstam, Martin et al. (eds.) (1996), *Proceedings of Euralex-'96*. Goetheborg University: Department of Swedish, 2: 715-28.

Didattica e risorse digitali

Appunti per un percorso sul tema 'tempo'

Giancarlo Scarpa

(Università Ca' Foscari Venezia, Italia)

Abstract This paper aims at showing how some phrases, in classical as well as in modern languages, which are found in ancient Greek and Latin inscriptions as well as literary epigrams, present a considerable continuity until our time. This can be particularly appreciated by using a digital search tool recently developed: the *Memorata poetis* thematic search engine. In this respect, a case study is investigated, pivoting on the 'Time' theme, and diachronic intertextual search outputs are offered. All results are accompanied by short comments, relying on the consistency of a widespread poetic memory including texts up to Proust, Montale, through Shakespeare and others. Their number may be considerably improved by the implementation of the *Memorata poetis* search engine.

Keywords Greek inscriptions. Latin inscriptions. Literary epigrams. The Time theme. Poetic memory.

Riprenderò a parlare in questo Convegno proprio dal punto in cui mi ero interrotto negli anni scorsi;¹ ma il progettato database *Memorata poetis* era allora quasi tutto nelle menti degli ideatori, laddove ora, grazie alla fatica di numerosi operatori delle marcature tematiche, ha ormai raggiunto livelli soddisfacenti di ampiezza, di efficienza e dunque di validità – almeno per la parte non piccola che riguarda la coppia di *Anthologiae* (*Graeca* e *Latina*) di cui ci occuperemo qui; però ancora mancano altri testi a esse immediatamente avvicinati, come per esempio le epigrafi sepolcrali in versi raccolte insieme da Werner Peek.² Una scelta piuttosto casuale di temi presi dalla *lanx exquisitis rebus satura* sembrava la via migliore per mostrare l'utilità del lavoro svolto da tante persone: progettista del software, operatori informatici, ricercatori addetti alla 'catalogazione'; la preferenza ultima è però stata quella di offrire un percorso che salisse dall'ipotesto al testo, con l'intento di mostrare ancora una volta come in fasi assai diverse della storia occidentale le eterne tematiche umane

1 Una prima presentazione dello strumento, sul cui impianto si basa *Anthologiae*, e i suoi modi d'impiego, è stata offerta da chi scrive: Scarpa 2014.

2 Le sue straordinarie raccolte di iscrizioni in versi, che amo definire la *Antologia di Spon River* del mondo greco antico, sono edite da Ares Publ., Chicago 1988 (cf. Peek 1955).

che attraversano le antologie antiche rivivessero per sempre grazie alla continuità della memoria poetico-letteraria. Il campionario non apparirà forse abbastanza esteso, né l'analisi approfondita o sottile quanto la causa meritava, e di ciò mi scuso in anticipo; il fine del mio contributo era tuttavia modesto, cioè preparare un'unità didattica sopra un argomento ricorrente nello studio delle letterature, soffermandomi sui testi qui individuati, richiamando alla memoria un regesto di sensazioni personali accumulatosi in tanti anni di insegnamento nella scuola superiore: il tempo e la *morosa canities*. Sarebbe andato pure bene, a questo punto, un titolo come *morbus ipsa senectus*? (Ter., *Phorm.*, 575): ma è inutile cedere ai piagnistei, e sulla scia di Sen., *Epist.*, 49.4 mi limiterò a constatare che «non solebat mihi tam velox tempus videri; nunc incredibilis cursus apparet, sive quia admoveri lineas sentio, sive quia adtendere coepi et computare damnum meum».

Si dirà che un lavoro come quello che sto per esporre si è sempre fatto: certo, ma quante energie richiedeva finora preparare una unità didattica su un particolare tema? Non dimentico i lunghi pomeriggi trascorsi in biblioteca a compulsare – anche vanamente – *indices*, concordanze, lessici a stampa, per individuare luoghi utili a una ricerca, che quasi sempre si ampliava, con la necessità di tornare sopra le stesse pagine dei volumi rimessi sugli scaffali. Quanto più semplice (e sicura) oggi l'indagine con l'aiuto delle risorse digitali: in un'unica schermata si visualizza il menù dei rinvii ai testi da analizzare, con un clic si va a leggerli, grazie anche all'interscambio tra *Anthologiae* e *Musisque Deoque*, o gli altri archivi di *Poeti d'Italia in lingua latina*, magari controllandone le caratteristiche metriche su *Pede certo*.

Al tema della brevità, precarietà, transitorietà della vita, cioè del 'sentimento' del tempo, la poesia occidentale è da sempre interessata: basterà richiamare le foglie di Omero (*Il.*, 6, 119-236), ispiratrici in seguito del grido, composto ma dolente, di Mimnermo (fr. 2 W) e molti altri ancora. Dalle foglie ai fiori: l'effimera rosa, così presente nella produzione letteraria di ogni età, diviene simbolo di amore, vita e morte nel contempo, come ricorderà anche Orazio (*car.*, 2, 13, 13-14 «nimis brevis flores amoenae... rosae»). Sciocco chi si crede 'eterno': il domani è del tutto incerto, perciò occorre cercare di godere dell'attimo, pur nella consapevolezza che anche questo dura pochissimo:³ rendendo ancor più drammatico il nostro vivere, perché è solo questo l'istante che 'strappiamo' al nostro signore – il Tempo, appunto – e di cui possiamo dirci noi pienamente padroni.

Finora ci siamo illusi pensando di saper governare il tempo, assegnandogli il compito di agire con espressioni quali *hora fugit*, *tempus volat* e così via; queste le 'certezze' che, messe in conto, ci fanno credere di avere tra

3 Quattro milionesimi di secondo; questa la durata che oggi la scienza, con l'ausilio della PET, ha assegnato alla percezione della realtà. Si veda Livio 2005, e gli allegati al *Corriere della Sera*, «Corriere Salute» di giugno e luglio 2005.

le mani il suo uso; in realtà, come insegna Petrarca,⁴ proprio l'attribuire al tempo azioni da noi percettibili mette a nudo tutta la rassicurante inadeguatezza del vocabolario umano (o forse del *logos* stesso) che neppure sa esprimere i propri sentimenti. Ciò che resta è una serena, pacata, virile certezza della fine, che consiglia al saggio la coscienza dei propri limiti, la naturale disposizione al piacere, l'ovvio invito a godersi la vita connesso al motivo del tempo inesorabile e al tema del *carpe diem*.

Per mettere alla prova la biblioteca ben collaudata di *Memorata poetis* insieme a quella più recente di *Anthologiae*, farò uso di pochi esempi attinenti a *tempus edax, tempus fugiens et irreparabile*, seguendone la persistenza dall'antichità a oggi, nella produzione letteraria italiana ed europea. Ribadendone gli scopi didattici, ho strutturato il discorso 'da generale a particolare', limitandomi a presentare i testi scelti quasi senza discussione o approfondimento, nella fiducia che la poesia della memoria basti da sola a risvegliare la memoria poetica in chi ascolta o legge.

Dunque, la 'sofferenza' del tempo che è in noi e ci sovrasta, dal quale paradossalmente percepiamo di dipendere nel momento stesso in cui lo amministriamo, fu avvertita dagli uomini presumibilmente sin da quando ebbero la necessità di collocarsi nell'universo e cominciarono a pensare al proprio essere: distinguendo un tempo ciclico, in cui inizio e fine coincidono (l'anno, il mese, il giorno) e uno lineare, visto come una sequenza ordinata di istanti. I Greci mostrano di avere almeno due concetti fondamentali di tempo, espressi da *χρόνος* e *αἶών*: il tempo numerato, colto nella sua successione di attimi (*Χρόνος* 'invecchia': Aesch., *Prom.*, 981), e il tempo-durata, colto nella totalità dell'insieme, quello che viene detto il tempo vitale (cf. Cresci 2000, 36 ss). Significativi risultano due passi, uno di Platone, l'altro di Aristotele, da cui si evince quanto appena detto; il primo (Plat., *Tim.*, 37): «ordinando il cielo crea dell'eternità che rimane nell'unità una immagine eterna che procede secondo il numero, quella che abbiamo chiamato tempo [*χρόνος*] [...] il divenire e il movimento della realtà si ritmano secondo un ordine e una loro misurabile durata, che è il tempo, immagine mobile dell'eternità; [...] poiché idealmente il tempo è l'unità della durata, esso non esisterebbe se non si scandisse in un ritmo di passato-presente-futuro, segnato dal moto (misurabile) degli astri. Misurabile, dunque, il mondo corporeo, e riducibile a punti e a linee il tempo»; ecco l'altro (Aristot., *Phys.*, 4, 11) «tempo dunque è il *numero*

4 Petrarca, *sen.*, 12.1: «io considero che velocissima corre, anzi vola la vita. Volano gli anni, diceva Cicerone, ed io di buon grado userei, se vi fosse, qualche parola, che più del volare esprimesse il rapido dileguarsi del tempo. Ma lento è il volo d'ogni augello ragguagliato al volare de' giorni nostri. Chè di quelli ci vien fatto distinguere il moto delle ali, l'avanzarsi che fan nello spazio, e l'avvicinarsi al termine, ed il raggiungerlo; ma a noi medesimi, tranne alcuni pochissimi privilegiati dal cielo per singolari doti d'ingegno e di cuore, passa il tempo senz'avvedercene, e inaspettata improvvisa ci si para innanzi la mèta...» (in Fracassetti 1869).

(la misura) del movimento secondo il prima e il poi [τοῦτο γὰρ ἔστιν ὁ χρόνος, ἀριθμὸς κινήσεως κατὰ τὸ πρότερον καὶ ὕστερον]»; ma poiché il prima e il poi non esistono se non nella coscienza, il tempo è, appunto, l'unità di coscienza del movimento, l'unità del passato, del presente e del futuro, onde il tempo sta alla coscienza come la continuità sta al luogo). In tale visione, abbiamo appena detto, l'inizio e la fine si congiungono per 'ricominciare': è il tempo ciclico, che già i primi 'filosofi' o, meglio, i 'fisiologi' esprimevano. Ma esiste il tempo di per sé? Lasciamo che a rispondere sia il poeta epicureo – pur sapendo che opinione diametralmente opposta hanno gli stoici – perché mi pare più coerente con quanto dirò (Lucr., 1, 459-463, traduzione di Luca Canali).

Tempus item per se non est, sed rebus ab ipsis consequitur sensus, transactum quid sit in aevo, tum quae res instet, quid porro deinde sequatur: nec per se quemquam tempus sentire fatendumst semotum ab rerum motu placidaque quiete.

Ugualmente **il tempo non esiste di per sé**, ma dalle cose stesse deriva il senso di ciò che è trascorso nei secoli, di ciò che incombe, o poi seguirà nel futuro. **Né si deve ammettere che alcuno avverta il tempo separato dal movimento** delle cose e dalla placida quiete.

Si noti come Lucrezio (cioè Epicuro) segua Aristotele nel ribadire l'importanza della cinesi in connessione con il sentimento del tempo; e inoltre si ponga mente all'impossibilità di avvertire l'azione del nostro 'tiranno', dunque di definirla. Petrarca arriverà a questa conclusione in totale indipendenza dal testo lucreziano, poiché il *de rerum natura* comincerà a circolare in Italia solo dopo la fortunata 'scoperta' di Poggio Bracciolini a Murbach, intorno al 1417. Ma cos'è allora il tempo? Agostino aveva risposto:

Se nessuno me lo chiede, lo so; se dovessi spiegarlo a chi me ne chiede, non lo so: eppure posso affermare con sicurezza di sapere che se nulla passasse, non esisterebbe un passato; se nulla sopraggiungesse, non vi sarebbe un futuro: se nulla esistesse, non vi sarebbe un presente. Passato e futuro: ma codesti due tempi, in che senso esistono, dal momento che il passato non esiste più, che il futuro non esiste ancora? E il presente, alla sua volta, se rimanesse sempre presente e non tramontasse nel passato, non sarebbe tempo, ma eternità. Se dunque il presente, perché sia tempo, deve tramontare nel passato, in che senso si può dire che esiste, se sua condizione all'esistenza è quella di cessare dall'esistere; se cioè non possiamo dire che in tanto il tempo esiste in quanto tende a non esistere?... Un fatto è ora limpido e chiaro: né futuro né passato esistono. È inesatto dire che i tempi sono tre: passato, presente e futuro. Forse sarebbe esatto dire che i tempi sono tre: presente del passato, presente del presente, presente del futuro. Queste tre specie di tempo esistono in qualche modo nell'animo e non vedo altrove: il presente del passato è memoria, il presente del presente la visione, il presente del futuro l'attesa. (Aug., *Conf.*, 14, 20; la traduzione è di Carlo Carena)

Sorvolerò qui su altre importanti pagine contenenti le riflessioni di pensatori quali Seneca e lo stesso Agostino, mi limiterò a parlare dell'esperienza fatta sugli archivi poetici di *Anthologiae* onde ricavare da una produzione in versi trascurata e 'minore' qualche tessera utile al nostro discorso. Do per scontato che i testi epigrammatici ed epigrafici possiedano uguale dignità e valore letterario,⁵ dunque invito a leggere le iscrizioni presenti nella raccolta di Peek intrecciate senza discriminazioni con i pezzi della *Anthologia Graeca* arrivati sino a noi per via di codici manoscritti medievali.

Aprondo il database *Anthologiae* e cliccando sul nodo/ramo principale 'Tempus' dall'albero/tema/macro-area 'Homines' – tralascio la stessa voce, inserita come foglia nel ramo 'Philosophia' – appare la seguente schermata:

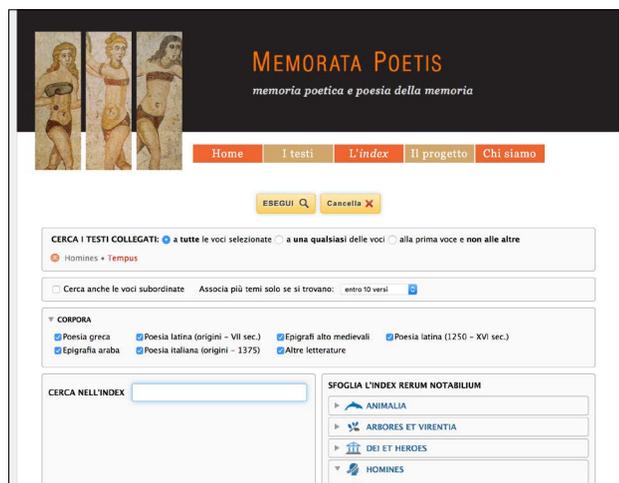


Figura 1. Schermata del percorso di ricerca sul sito <http://www.memoratapoetis.it/public/> (2017-11-28)

Ho scelto e sistemato nella sequenza utile al percorso che intendo qui svolgere i seguenti testi.

5 Contro le opinioni di alcuni: Scarpa 2014, 22.

1. Leonida, AG, 7, 472 + 472b (III secolo a.C.)⁶

Μυρίος ἦν, ὠνθρωπε, χρόνος πρὸ τοῦ, ἄχρι πρὸς ἧῶ
ἦλθες, χῶ λοιπὸς **μυρίος** εἰς Αἴδην.
τίς μοῖρα **ζωῆς** ὑπολείπεται, ἦ ὅσον ὅσσον
στιγμῆ καὶ **στιγμῆς** εἶ τι χαμηλότερον;
μικρῆ σευ **ζωῆ** τεθλιμμένη· οὐδὲ γὰρ αὐτῆ
ἦδεῖ, ἀλλ' **ἐχθροῦ** **στυγνοτέρη** **θανάτου**.
ἐκ τοίης ὠνθρωποὶ ἀπηκριβωμένοι **ὀσῶν**
ἀρμονίης **ὑψοῦντ'** ἠέρα καὶ νεφέλας·
ὠνερ, ἴδ', ὡς ἀχρεῖον, ἐπεὶ περὶ νήματος ἄκρον
εὐλῆ ἀκέρκιστον λῶπος ἐφεζομένη·
τοῖον τὸ **ψαλάθρειον ἀπεψιλωμένον** οἶον
πολλῶ ἀραχναίου στυγνότερον σκελετοῦ.
ἦοῦν ἐξ ἠοῦς ὅσσον σθένος, ὠνερ, ἐρευνῶν
εἰς **ἐν λειτῆ** **κεκλιμένος βιοτῆ**·
αἰὲν τοῦτο νόω **μεμνημένος**, ἀχρις ὀμιλῆς
ζωοῖς, ἐξ οἴης ἠρμόνισαι **καλάμη**.

AG, 472b

Χειμέριον ζωὴν ὑπαλεύσο, νεῖο δ' ἐς ὄρμον,
ὡς κῆγῶ Φεῖδων ὀ Κρίτου, εἰς Αἴδην.

Infinito, prima di nascere, il tempo; **infinito**,
uomo, quello che resta, nell'Averno.
Quale porzione di vita rimane, se non come un **punto**
O meno, se possibile, d'un **punto**?
Corta la vita e oppressa; per sé non ha nulla di dolce,
ma odiosa è più della funesta morte.
Uomini, tale è la vostra compagine d'ossa, e nell'aria
Voi **vi librate**, fra le nubi, in alto!
Che stupidaggine, uomo! C'è il verme all'estremo del filo,
su quel mantello non tocco da spola.
Tale il **nudo teschio spelato di chiome**: un oggetto
Odioso più che scheletrito ragno.
Uomo, **giorno da giorno** con tutte le forze scrutando,
d'una semplice vita resta pago!
Finché starai tra i viventi, **ricordi** sempre la mente
Di che tenue compagine sei fatto.

Fuggi bufere di vita, dirigiti **al porto** (lo feci
io, Fidone di Crito) giù nell'Ade.

Appunti sul testo

1. Nell'epigramma si segnala, sul piano stilistico, il ricorrere della anafora, figura retorica impiegata spesso in contesti simili a questo.
2. Infinito (vv. 1-2): **μυρίος** è vocabolo che rinvia al numero, definendo **χρόνος** come 'tempo numerato', concetto accostabile a quello di 'infinito' numerico e spaziale (spazio-movimento-tempo).
3. Un punto (mi sovviene l'uomo definito 'atomo vivente') al v. 4. Si richiama per questo il pensiero di Luigi Capuana e il lavoro di E. Comoy Fusaro, *Forme e figure dell'alterità. Studi su De Amicis, Capuana e Camillo Boito*, Ravenna 2009, 91 s.
4. La visione della vita (vv. 5-6).
5. Lessico 'macabro' (vv. 7, 9, 11, 12).
6. Richiamo alla miseria e alla vanità umana (v. 8), ma vedi anche AG, 12, 32; 11, 289.
7. Invito a 'godere' la vita, una sorta di *carpe diem*.
8. Il motivo del ricordo.
9. Al v. 14 si coglie l'invito a una vita fondata sulla *medietas*.
10. Al v. 16 c'è un intrigante richiamo a questa 'canna' (**καλάμη**) che potrebbe diventare 'pensante' per il nostro Pascal; traducendo 'tenue compagine' si perde la suggestione.
11. La morte come 'agognato porto' (472b).

6 La traduzione (come per quasi tutti gli altri testi) è di Pontani 1978-1981.

La prima tentazione, di fronte al testo, sarebbe quella di definire l'autore un poeta 'scapigliato', o forse 'maledetto', per il lessico e il realismo macabro, che rimanda pure a figure d'artista del nostro Seicento, quali Ciro di Pers (*Della miseria e vanità umana*, 194-99):

Goda pur lieto, goda
tra le festose cene
i delicati cibi,
in fin ch'egli divenga
di mille immobil vermi
putrido cibo, abbominevol esca...

O ancora il sonetto *Miseria umana*:

L'uomo, che sì poche de la vita ha l'ore,
e ne conta a fatica una gioconda,
è di sospir, di pianto, un'aura, un'onda:
piangendo nasce⁷ e sospirando mòre.
Aura ch'avviva un inonesto ardore,
onda che sprezza una discreta sponda;
aura che scuote **una caduta fronda**,⁸
onda che irriga un momentaneo fiore.
E, benché aspiri a sempiterno vanto,
per quelle vie che strepitose corse
a pena un lieve suon mormora alquanto.
Mentre l'uomo formò, Prometeo forse
il duro fango distemperò col pianto,
e co' sospir lo spirito gli porse.

2. Pallada, AG, 15, 20 (fine IV secolo d.C.)

Σιγῶν παρέρχου τὸν **ταλαίπωρον βίον**
αὐτὸν σιωπῇ τὸν χρόνον μιμούμενος;
λαθῶν δὲ καὶ **βίωσον**, εἰ δὲ μή, θανών.

Silente passa questa vita misera e in questo
gran silenzio al tempo adeguati:
nascosto vivi o, se non puoi, cadavere.

7 Superfluo ricordare come il tema del pianto alla nascita ricorra in Lucrezio, Seneca, Leopardi, tanto per limitarsi a 'grandi' autori a tutti noti.

8 Con una insignificante variazione rispetto a questo, il tema della foglia staccata dal ramo e trasportata dal vento appare in Arnault, *La feuille*, poesia che viene tradotta e fatta propria da Leopardi (*Lungi dal proprio ramo*, c. 35).

Appunti sul testo

1. Il trascorrere del tempo non fa rumore (così Leopardi, *Canto notturno*, v. 72 «tacito infinito andar del tempo»: memoria poetica del Recanatese?).
2. Si è tradotta qui l'espressione 'imitare il tempo' (τὸν χρόνον μιμούμενος) con il verbo 'adeguarsi', che ne è l'interpretazione.
3. Precetto epicureo del 'vivi nascosto' (λάθε βιώσας): con variante dovuta alla 'memoria' dell'epigrammista?

3. L'azione del tempo è ben richiamata in: Platone, *AG*, 5, 80 (427-347 a.C)

Μῆλον ἐγὼ· **βάλλει** με φίλων σέ τις· ἀλλ' ἐπίνευσον, **Ξανθίππη**· κἀγὼ καὶ σὺ **μαραίνόμεθα**. Sono una mela: mi scaglia taluno che t'ama, Santippe, digli di sì: **sfiorisci**, come me.

Appunti sul testo

1. Rinvio al mito.
2. Il verbo greco μαραίνω esprime una lenta e progressiva consunzione; si veda tra altri esempi *Soph., Ai.*, 714: πάνθ' ὁ μέγας χρόνος μαραίνει.

4. Ancora Platone, *AG*, 9, 51 con un distico la riassume bene

Αἰὼν πάντα φέρει· δολιχὸς χρόνος οἶδεν ἀμείβειν **Tutto il tempo si porta: traveste**, una **lunga durata**, οὐνομα καὶ μορφήν καὶ φύσιν ἠδὲ τύχην. nome e figura e indole e fortuna.

Appunti sul testo

1. Si fa notare come la traduzione di Filippo Maria Pontani rinvii ai *Sepolcri* del Foscolo («le reliquie | della terra e del ciel traveste il tempo», vv. 21-2): voluta citazione per tradurre il verbo ἀμείβειν, o memoria poetica operante nel traduttore?
2. L'aggettivo δολιχὸς richiama una corsa piuttosto lunga.
3. Nota come sia usato αἰών (il tempo vitale) per indicare la prima azione e χρόνος (il tempo numerato) per esprimere la durata.

5. Rufino, AG, 5, 76 (II secolo d.C. ?)

Αὕτη πρόσθεν ἔην ἑρατόχροος, εἰρόμασθος,
εὖσφυρος, εὐμήκης, εὖοφρυς, εὐπλόκαμος,
ἡλλάχθη δὲ **χρόνω** καὶ γήραι καὶ πολιαίσι,
καὶ νῦν τῶν προτέρων οὐδ' ὄναρ οὐδὲν ἔχει·
ἄλλοτρίας δὲ τρίχας καὶ ῥυσῶδες <τὸ> πρόσωπον,
οἶον γηράσας οὐδὲ πίθηκος ἔχει.

Ebbe una volta una tinta leggiadra, freschezza di seni,
caviglie, taglia, ciglia, chiome belle.

L'hanno mutata l'età, la vecchiezza, le chiome canute;
non ha più l'ombra del fascino antico,
piena di frezze posticce, di **rughe** che coprono il viso:
tante una scimmia vecchia non ne mostra.

Appunti sul testo

1. L'azione del tempo si percepisce nelle cose e nelle persone (rughe, etc.).
2. Richiama la miseria e la vanità umana.
3. Si vedano alcuni sonetti di Shakespeare (testo e traduzione in <http://www.shakespeareweb.it>).

II

Quando quaranta inverni avranno aggredito la tua fronte
e scavato fonde trincee nel campo della tua bellezza,
la superba veste della tua gioventù or tanto ammirata,
sarà considerata un cencio di nessun valore:
se allora ti venisse chiesto dove giace il tuo fascino
e dove si è perso l'amore dei tuoi ruggenti giorni,
ammettere che è in fondo ai tuoi occhi incavati
sarebbe penosa vergogna ed inutile vanto

When forty winters shall beseige thy brow,
And dig deep trenches in thy beauty's field,
Thy youth's proud livery, so gazed on now,
Will be a tatter'd weed, of small worth held:
Then being ask'd where all thy beauty lies,
Where all the treasure of thy lusty days,
To say, within thine own deep-sunken eyes,
Were an all-eating shame and thriftless praise

III

Guardati allo specchio e di' al volto che vedi
che è ormai tempo per quel viso di crearne un altro;

Look in thy glass, and tell the face thou viewest
Now is the time that face should form another;

XII

Quando seguo l'ora che batte il passar del tempo
e vedo il luminoso giorno spento nella tetra notte,
quando scorgo la viola ormai priva di vita
e riccioli neri striati di bianco,
quando vedo privi di foglie gli alberi maestosi
che un dì protessero il gregge dal caldo
e l'erbe d'estate imprigionate in covoni
portate su carri irte di bianchi ed ispidi rovi,
allor, pensando alla tua bellezza, dubbio m'assale
che anche tu te ne andrai tra i resti del tempo,
perché grazie e bellezze si staccan dalla vita
e muoiono al rifiorir di altre primavere:
e nulla potrà salvarsi dalla lama del Tempo
se non un figlio che lo sfidi quand'ei ti falcerà.

When I do count the clock that tells the time,
And see the brave day sunk in hideous night;
When I behold the violet past prime,
And sable curls all silver'd o'er with white;
When lofty trees I see barren of leaves
Which erst from heat did canopy the herd,
And summer's green all girded up in sheaves
Borne on the bier with white and bristly beard,
Then of thy beauty do I question make,
That thou among the wastes of time must go,
Since sweets and beauties do themselves forsake
And die as fast as they see others grow;
And nothing 'gainst Time's scythe can make defence
Save breed, to brave him when he takes thee hence.

XXII

Lo specchio non mi convincerà che sono vecchio,
finché tu e giovinezza avrete la stessa età;

My glass shall not persuade me I am old,
So long as youth and thou are of one date;

LXV

O come potrà reggere il fresco alito d'estate
alla rovinosa stretta dei martellanti giorni,
se rocche invulnerabili non sono tanto solide
né porte d'acciaio salde al rovinar del Tempo?
O tragico pensiero! Dove, ahimè, nascondere
al forziere del Tempo il suo più bel gioiello?
Qual possente mano potrà frenarne il lesto passo?
O chi saprà vietargli lo sterminio di bellezza?

O, how shall summer's honey breath hold out
Against the wreckful siege of battering days,
When rocks impregnable are not so stout,
Nor gates of steel so strong, but Time decays?
O fearful meditation! Where, alack,
Shall Time's best jewel from Time's chest lie hid?
Or what strong hand can hold his swift foot back?
Or who his spoil of beauty can forbid?

LXXVII

Lo specchio ti dirà come si logori la tua bellezza,
la meridiana come si consumi il tuo prezioso tempo;
le pagine vuote imprimeranno i tuoi pensieri
e dal libro potrai trarre questo insegnamento:
le rughe che il tuo specchio rifletterà sincero
susciteranno in te il ricordo di tombe spalancate;
dal segreto muoversi dell'ombra sulla meridiana,
vedrai il progredir furtivo del tempo all'eternità [...]

Thy glass will show thee how thy beauties wear,
Thy dial how thy precious minutes waste;
The vacant leaves thy mind's imprint will bear,
And of this book this learning mayst thou taste.
The wrinkles which thy glass will truly show
Of mouthed graves will give thee memory;
Thou by thy dial's shady stealth mayst know
Time's thievish progress to eternity [...]

CXV

Ma pensando al tempo, i cui infiniti eventi
s'insinuan tra giuramenti e mutan decreti di re,
spogliano sacre bellezze, stroncan audaci disegni
volgono all'incostanza le più salde menti:
ahimè – temendo la tirannia del Tempo – perché
non dissi allora: “T'amo d'un amor supremo”,
quando ero certo al di là di ogni incertezza
di consacrare quel presente, ignaro del domani?

But reckoning time, whose million'd accidents
Creep in 'twixt vows and change decrees of kings,
Tan sacred beauty, blunt the sharp'st intents,
Divert strong minds to the course of altering things;
Alas, why, fearing of time's tyranny,
Might I not then say “Now I love you best,”
When I was certain o'er uncertainty,
Crowning the present, doubting of the rest?

CXXIII

No Tempo, mai ti vanterai ch'io cambi;
 le piramidi che innalzi con sempre nuova possa
 non mi dicon niente, non han niente di nuovo:
 non son che nuove vesti di cose già vedute.
 È breve l'arco della vita, perciò guardiam stupiti
 il vecchio che ci imponi come fosse nuovo,
 e che vogliamo credere fatto a nostro gusto
 piuttosto di pensare che già ne udimmo dire.
 Io ti sfido Tempo e sfido i tuoi registri,
 perché non mi sorprende il tuo presente od il passato
 le tue vestigia mentono e mente quanto vediamo
 fatto grande o piccolo dalla tua continua furia:
 questo io ti giuro e questo manterrò,
 a scorno tuo e della tua falce, io non cambierò.

No, Time, thou shalt not boast that I do change:
 Thy pyramids built up with newer might
 To me are nothing novel, nothing strange;
 They are but dressings of a former sight.
 Our dates are brief, and therefore we admire
 What thou dost foist upon us that is old,
 And rather make them born to our desire
 Than think that we before have heard them told.
 Thy registers and thee I both defy,
 Not wondering at the present nor the past,
 For thy records and what we see doth lie,
 Made more or less by thy continual haste.
 This I do vow and this shall ever be;
 I will be true, despite thy scythe and thee.

6. Sullo stesso tema Giuliano Egizio, *AG*, 6, 18 (VI secolo d.C.) rinvia al motivo dello specchio, su cui, come abbiamo appena visto, torna più volte Shakespeare nei sonetti, intrisi di espressioni piuttosto singolari a proposito dell'azione del Tempo.

Λαῖς ἀμαλδυνθεῖσα χρόνῳ περικαλλέα μορφήν
 γηραλέων στυγέει μαρτυρίην ῥυτίδων
 ἔνθεν πικρὸν ἔλεγχον ἀπεχθήρασα κατόπτρου
 ἀνθετο δεσποίνῃ τῆς πάρος ἀγλαίης.
 «Ἄλλὰ σύ μοι, Κυθέρεια, δέχου νεότητος ἐταῖρον
 δίσκον, ἐπεὶ μορφή σὴ χρόνον οὐ τρομέει».

Laide, **gualcita** la bella figura dal tempo, detesta
 I testimoni di senili rughe.
 Ecco aborre lo **specchio** – denuncia che punge –, lo dona
 a chi protesse quel fulgore antico:
 «Cipride, a te questo disco, degli anni migliori compagno:
 la tua figura non lo teme, il tempo».

Appunti sul testo

1. Il significato di ἀμαλδυνθεῖσα è più forte di quello assegnatogli dal traduttore; in effetti il verbo indica l'azione dell'annientamento.

7. Ritroviamo il tentativo di definire il tempo e la sua azione in *AL* 676

Omnia tempus agit, cum tempore cuncta trahuntur.
 Alternant elementa uices et tempora mutant.
 [...]

Tempus et hora uolat, momentis **labitur aetas**.
 Omnia dat tollit minuitque uolatile tempus.
 Ver aestas autumnus hiems: redit annus in annum.
 Omnia cum redeant, homini sua non redit aetas.

Il tempo tutto porta, tutto con il tempo se ne va.
 Gli elementi cambiano alternativamente e mutano le stagioni.
 [...]

Il tempo e l'ora volano, la vita scorre via in un attimo.
 Il tempo in volo dà, toglie, riduce tutto.
 Primavera, estate, autunno, inverno: l'anno ritorna sempre.
Tutto ritorna, ma per l'uomo **non ritorna la vita**.

Appunti sul testo

1. Cf. *infra* 9 per il verbo *labi*.

8. 'Scendendo' nei sottotemi di 'Tempus', troviamo il 'Tempus edax' che rinvia a: Anonimo, *AG*, 7, 225

Ψήχει καὶ πέτρην ὁ πολὺς χρόνος οὐδὲ σιδήρου
φείδεται, ἀλλὰ μὴ πάντ' ὀλέκει **δρεπάνη**
ὡς καὶ Λαέρταο τόδ' ἠρίον, ὃ σχεδὸν ἀκτᾶς
βαῖον ἄπο, ψυχρῶν λείβεται ἔξ ὑετῶν.
οὖνομα μὴν ἥρωος ἀεὶ νέον· οὐ **γὰρ αἰοῖδ' ἄς**
ἀμβλύνειν αἰῶν, κῆν ἐθέλη, δύναται.

Rode il tempo diuturno la pietra, né ferro risparmiava:
con una **falce** sola tutto stronca.
Ecco la tomba, qui, di Laerte: vicino alla riva,
ha per libame un po' di pioggia fredda.
Ma dell'eroe vigoreggia, per sempre giovane, il nome:
non fiacca i canti, se pur vuole, **il tempo**.

Appunti sul testo

1. Immagine della morte armata di falce: si identifica con il tempo che agisce, consuma e porta la vita alla fine – sulla quale però trionfa la poesia eternatrice (vv. 5-6); si veda *supra* il sonetto CXXIII di Shakespeare.

9. Ovidio, *AL*, 269 (43 a.C-18 d.C)

De aetate

Vtendum est aetate; cito pede labitur aetas
Nec bona tam sequitur quam bona prima fuit.
Heu me nunc miserum! laxantur corpora **rugis**
Et perit in nitido qui fuit ore color.

Bisogna far uso della vita; velocemente scorre via
e non prosegue tanto lieta, quanto lo fu all'inizio.
Ahimé ora infelice! Le carni si rilassano per le **rughe**
e **scompare quello che fu il colore in uno splendido viso.**

Appunti sul testo

1. Labi: 'scivolare', è verbo che piace all'Orazio della celebre ode 2.14 (*Eheu fugaces, Postume, Postume, labuntur anni*); cf. *supra* 7.
2. Rinvio al motivo delle rughe (*supra* 6).

10. Seneca, *Epigr.*, 1 (problema dell'attribuzione?)

Omnia **tempus edax** depascitur, omnia **carpit**,
Omnia sede mouet, nil sinit esse diu.
Flumina deficiunt, profugum mare litora siccata,
Subsidunt montes et iuga celsa ruunt.
Quid tam parua loquor? moles pulcherrima caeli
Ardebit flammis tota repente suis.
Omnia mors poscit. Lex est, non poena, perire;
Hic aliquo mundus tempore nullus erit.

Il tempo ingordo tutto divora, tutto **strappa via**,
tutto smuove dalla sua sede, nulla lascia esistere a lungo,
i fiumi si disseccano, il mare ritraendosi prosciuga i lidi,
i monti franano, gli alti gioghi crollano.
Perché parlo di piccole cose? La bellissima volta del cielo
All'improvviso brucerà tutta per le sue fiamme.
La morte tutto richiede. È legge, non pena, morire:
un giorno non esisterà più questo mondo.

Appunti sul testo

1. Il *tempus edax*, 'tempo ingordo' rinvia al 'tempo divoratore' del sonetto XIX di Shakespeare.

2. Il verbo *carpere*, 'strappare via', indica l'azione che si compie quando si coglie un fiore e rinvia al *carpe diem* di oraziana memoria (cf. per questo Paoli 1969, 53 n. 8).
3. Si noti il colore stoico del contenuto.

11. Giuliano Egizio, AG, 6, 19 (VI secolo d.C.)

Κάλλος μὲν, Κυθέρεια, χαρίζεται, ἀλλὰ **μαραίνει** Tu la bellezza, Ciprigna, regali, ma **tacito striscia**,
ὁ χρόνος ἐρπύζων σὴν, βασιλεια, χάριν· il tuo regalo **disfiorendo, il tempo.**
 δῶρου δ' ὑμετέροιο παραπταμένου με, Κυθήρη, Lunghi il dono volò, Citerea: del dono ricevi,
 δέχυσσο καὶ δῶρου, πότνια, μαρτυρίην. regina augusta, questo testimone.

Appunti sul testo

1. Richiama l'azione silenziosa del tempo (come abbiamo notato anche *supra* 2).
2. Ancora il verbo *μαραίνω* (come *supra* 3); ricorrendo al verbo 'disfioreare' il traduttore sta pensando alla vita come morte continua.

12. Epigono di Tessalonica, AG, 9, 261 (?)

Ἦ πάρος εὐπετάλοισιν ἐν οἰνάνθαις νεάσσασσι Giovane un tempo, ravvolta di floridi pampini, diedi
 καὶ τετανῶν βοτρύων ῥᾶγα κομισσαμένη Grappoli densi di succosi grani.
 νῦν οὕτω γραιοῦμαι. ἴδ', ὁ χρόνος οἷα δαμάζει· Ora invecchio, lo vedi. **Che cosa non domina il tempo!**
 καὶ σταφυλὴ γήρωσ αἰσθάνεται ῥυτίδων. Persino l'uva ha il senso delle rughe.

Appunti sul testo

1. Il tempo agisce anche sulle cose; a questo proposito si richiamerà la celebre lettera a Lucilio (12) dove Seneca descrive una visita alla sua vecchia casa di campagna; l'edificio, intaccato dai guasti del tempo, gli ricorda il proprio stato di essere effimero; e si veda subito *infra* AG, 9, 499.

13. AG, 9, 499 fa qui da cerniera al sottotema successivo: *Tempus fugiens et irreparabile*

Ἀργαλέως φέρεται πολὺς χρόνος, ἀλλὰ παρέρπων **Penosamente il tempo canuto procede**; strisciando
 καὶ φωνὰς **κλέπτει** φθεγγομένων μερόπων, **Ruba** la voce agli uomini parlanti,
 καὶ μὴ φαινόμενος τοὺς φαινόμενους **ἀφανίζει** quelli che appaiono fa, senza mai comparire, **sparire**,
 καὶ μὴ φαινόμενους εἰς φανερόν προφέρει. porta chi non appare ad apparire.
 ὦ ζωῆς **ἀόριστος** ἐν ἀνθρώποισι τελευτή, **Indefinito** fine di vita per gli uomini!
ἡμαρ ἐπ' ἡμαρ αἰεὶ πρὸς ζόφον ἐρχομένων. verso la notte, giorno dopo giorno.

Appunti sul testo

1. Lo 'strisciare' del tempo: anche *supra* 11.
2. La *morosa canities* si associa al tempo che trascorre, è passato e non tornerà.

14. A questa amara verità già Asclepiade, AG, 12, 50 (270 a.C.) aveva rivolto lo sguardo

Πῖν', Ἀσκληπιάδη. τί τὰ δάκρυα ταῦτα; τί πάσχεις;
οὐ σὲ μόνον χαλεπὴ Κύπρις ἐλήσιατο,
οὐδ' ἐπὶ σοὶ μούνω κατεθήξατο τόξα καὶ ἰοῦς
πικρὸς Ἔρως. **τί ζῶν ἐν σποδιῇ τίθεσαι;**
πίνωμεν Βάκχου ζῶρον πόμα· **δάκτυλος ἄώης.**
ἦ πάλι κοιμιστὰν λύχνον ἰδεῖν μένομεν;
πίνομεν· οὐ γὰρ ἔρως· μετὰ τοι χρόνον οὐκέτι πουλύν,
σχέτλιε, **τὴν μακρὰν νύκτ' ἀναπαυσόμεθα.**

Asclepiade, tracanna! Cos'è quel pianto? Che hai?
Preda non sei tu solo di Ciprigna,
né per te solo aguzza Cupido amaro gli strali.

Vivo e ridotto in cenere: perché?

Pretto liquore di Bacco beviamo, ché **il giorno è una spanna!**

O aspettiamo la torcia che ci addormenta?

Gagliardamente beviamo! Mio povero amico, fra poco

Riposeremo quella notte lunga.

Appunti sul testo

1. Rinvia naturalmente a AG XI 56 (*carpe diem*).
2. L'ultimo pentametro richiama Catullo, 5, 6 (*nox est nobis una dormienda*).
3. La brevità della vita e la visione senecana.
4. Vivere è sopravvivere alla propria morte.

15. All'uomo che in Lucrezio e Seneca nasce piangendo (tema pessimistico, presente tra gli altri in Pallada, AG, 10, 84, in Ciro di Pers e ancora in Leopardi, *Canto notturno*, 39 ss.), come ai chicchi di caffè di trilussiana memoria, si prospetta il baratro del nulla sin dal primo schiudersi della vita.

16. Ammiano, AG, 11, 13 (II secolo d.C.)

Ἦώς ἐξ ἠοῦς παραπέμπεται, εἴτ' ἀμελούντων
ἡμῶν ἐξαίφνης ἦξει ὁ πορφύρεος
καὶ **τοὺς μὲν** τήξας, **τοὺς δ'** ὀπτήσας, **ἐνίους** δὲ
φυσήσας ἄξει πάντας ἐς ἐν **βάραθρον.**

Seguono aurore ad aurore, finché spensierati, d'un tratto
ci coglierà la livida parvenza:

gli **uni** struggendo, gli **altri** bruciando, **taluni** gonfiando,
tutti a un **baratro** solo porterà.

Appunti sul testo

1. L'espressione anaforica e poliptotica è presente in questo come in altri testi riferiti al tempo.

17. Pallada, AG, 10, 84 (fine IV secolo d.C)

Δακρυχέων γενόμην καὶ **δακρύσας** ἀποθνήσκω·
δάκρυσι δ' ἐν πολλοῖς τὸν βίον εὖρον ὄλον.
ὦ **γένος** ἀνθρώπων **πολυδάκρυτον**, ἀσθενές, οἰκτρόν,
συρόμενον κατὰ γῆς καὶ διαλυόμενον.

Lacrime piansi nascendo, così **lacrimando** perisco:
tutta fu tra le lacrime la vita.
Ahi, **lacrimata stirpe** degli uomini, grama, penosa!
Trascinata sotterra, si dissolve.

Appunti sul testo

1. Rinvia a Lucrezio, Seneca, Leopardi (*supra* 15).

Non manca però nell'immenso deposito della produzione letteraria classica – e prima ancora, ovviamente, nei sentimenti dell'animo umano – una 'reazione' a questo nemico-tempo.

Una foglia utile per la nostra riflessione si è quindi mostrata quella intitolata *Carpe diem*, che spunta dal ramo 'Vita'.

Qui è Asclepiade, *AG*, 12, 50 (si veda *supra* 14) che fa da cerniera: un invito al bere che viene da lontano, e sebbene emesso ad altri fini, approda in contesti come questi, che guardano al nostro tema.

18. Stratone, *AG*, 11, 19 (130 d.C.)

Καὶ **πίε** νῦν καὶ **ἔρα**, Δαμόκρατες· οὐ γὰρ ἐς αἰεὶ
πιόμεθ' οὐδ' **αἰεὶ** παῖσι συνεσσόμεθα.
καὶ στεφάνοις κεφαλᾶς πυκασώμεθα καὶ μυρίσωμεν
αὐτούς, πρὶν τύμβοις ταῦτα φέρειν ἑτέρους.
νῦν ἐν ἐμοὶ πιέτω μέθυσον τὸ πλέον ὅστέα τὰμά·
νεκρὰ δὲ Δευκαλίων αὐτὰ κατακλυσάτω.

Bevi, Damocrate, e ama! ché **non berremo per sempre**,
né coi ragazzi ci uniremo sempre.
Via, cingiamo corone, spalmiamoci tutti d'unguenti,
prima ch'altri ne adorni i nostri avelli.
Ora quest'ossa mie s'intridano tutte di vino,
Deucalione le inondi dopo morte!

19. Anonimo, *AG*, 11, 56

Πῖνε καὶ εὐφραίνου. τί γὰρ αὔριον ἢ τί τὸ μέλλον,
οὐδεὶς γινώσκει. μὴ τρέχε, μὴ κοπία·
ὡς δύνασαι, χάρισαι, μετάδος, φάγε, θνητὰ λογίζου·
τὸ ζῆν τοῦ μὴ ζῆν οὐδὲν ὅλως ἀπέχει.
πᾶς ὁ βίος τοιόσδε, **ῥοπή** μόνον· ἂν προλάβῃς, σοῦ,
ἂν δὲ θάνῃς, ἑτέρου πάντα, σὺ δ' οὐδὲν ἔχεις.

Bevi e gioisci! **Che cosa saranno il domani, il futuro, non si sa**. Niente furia, niente affanni!
Mangia, partecipa, dà come puoi, ragiona da uomo!
Tra il vivere e il non vivere c'è un filo.
Tutta la vita è una china. Per te, se l'afferri, ogni cosa;
se muori, è d'altri e nulla c'è per te.

Appunti sui testi

1. Un invito a godere e amare, facilmente diffuso a tutti i livelli e dunque entrato spesso nel repertorio dei detti popolari.
2. Per il motivo del bere, è scontato il rinvio ad Alceo e a Orazio (*carm.* 1.11; 1.37).
3. 'Definizione' della vita.
4. Vita e morte.
5. Tono 'scoptico' di alleggerimento della gravità del tema?

20. Stratone, AG, 12, 197 (130 d.C.)

«**Καιρὸν γνῶθι**» σοφῶν τῶν ἑπτὰ τις εἶπε, Φίλιππε·
πάντα γὰρ ἀκμάζοντ' ἔστιν ἔραστότερα·
καὶ σίκυος πρώτος που ἐπ' ἀνδῆροισιν ὄραθεῖς
τίμιος, εἶτα συῶν βρώμα πεπαινόμενος.

«**Cogli il momento**» fra i Sette sapienti taluno ci disse,
Filippo: sempre ciò che è in fiore piace.
Il cetriolo che prima tu scorgi sul margine vale,
poi si fa mezzo, è cibo per le scrofe.

Appunti sul testo

1. Anche nel contesto pederotico della *Musa puerilis* troviamo dunque un invito al *carpe diem* (dove il 'conosci te stesso' diviene un riconosci l'occasione).

21. Una spinta a cogliere i doni della giovinezza è anche in AL, 24

Amans amanti

Dic, quid agis, formosa Venus, si nescis amanti
Ferre uicem? perit omne decus, **dum deperit aetas**.
Marcent post florem uiolae, rosa perdit odorem,
Lilia post uernum posito candore **liquescunt**.
Haec metuas exempla precor, et semper amanti
Redde uicem, quia semper amat qui semper amatur.

Di', che fai, bella Venere, se non sai ricambiare
l'amante? Cade ogni bellezza, **mentre se ne va
la vita**. Le viole marciscono dopo esser fiorite,
la rosa perde il profumo, i gigli dopo l'inverno,
smesso il candore, **si putrefanno**. Ti prego,
temi questi esempi e ricambia sempre l'amante,
perché ama sempre chi sempre è amato.

Appunti sul testo

1. Rinvio a testi ben noti di autori del nostro Umanesimo (Lorenzo de' Medici) e Rinascimento.

22. Tutto quello che abbiamo elencato sinora, ricompare in AL, 378, dove si nota come nella cultura letteraria cristiana il lessico insista sull'eternità dell'anima, sulla vittoria della vita sulla morte corporale grazie alla rigenerazione del battesimo. E tuttavia il Tempo continua a definirsi mediante le vecchie espressioni topiche

Crede prius ueniens, Christi te fonte **renasci**:
Sic poteris mundus regna uidere dei.
Tinctus in hoc sacro mortem non sentiet umquam;
Semper enim uiuit, quem semel unda lauit.
Descende intrepidus: uitae fomenta **perennis**
Aeternos homines ista lauacra creant.
Ascende in caelos, animam qui in fonte lauisti,
Idque semel factum sit tibi **perpetuum**.
Peccato ardentis hoc fonte extinguite culpas.
Currite! quid statis? **tempus et hora fugit**.

Giungendo per la prima volta credi di **rinascere**
grazie alla fonte di Cristo: così purificato potrai
vedere i regni di Dio. Chi è immerso in questo
sacro fonte non percepirà mai la morte; infatti
vive sempre, chi l'onda lavò un'unica volta.
Scendi giù intrepido: codesti lavacri alimento
di vita **perenne** creano uomini **eterni**. Sali nei
cieli, anima che ti lavasti nel fonte e ciò che hai
fatto una sola volta sia per te **perpetuo**. Voi che
bruciate per il peccato, estinguette le colpe in
questo fonte. Correte! Che aspettate? **Il tempo e
l'ora fuggono**.

Appunti sul testo

1. Il Cristianesimo, portatore della speranza di un'altra vita nell'aldilà, ribalta i convincimenti ereditati dal mondo classico; ad esempio, la morte diventa la nascita, quindi 'vive per sempre' chi percorre le orme del Maestro, si affida alla Chiesa, ecc.
2. Un'espressione topica viene utilizzata senza evidenziare la gravità di ciò che significa, ma cogliendone la positività relativamente alla situazione presentata; si ricordi inoltre che *tempus volat, hora fugit* diventano etichette sugli orologi.

Attraverso gli archivi di *Memorata poetis* ci è consentito addentrarci entro spazi culturali di solito poco frequentati come i testi epigrafici del mondo arabo antico; qui ovviamente troviamo elementi sia di continuità che di separazione rispetto alle forme di pensiero e di espressione cui siamo più avvezzi nella nostra scuola. Che però il sentire dell'uomo di ogni dove, espresso nelle forme più comuni (come può essere un'epigrafe) possa facilmente intersecarsi con altre culture, lo dimostra quanto, a proposito del nostro tema, ha lasciato scritto Kahlil Gibran, un libanese trapiantato negli Stati Uniti nei primi decenni del Novecento:

E un astronomo disse: Maestro, parlatemi del Tempo.

E lui rispose:

Vorreste **misurare il tempo**, l'incommensurabile e l'immenso.

Vorreste regolare il vostro comportamento

e dirigere il corso del vostro spirito secondo le ore e le stagioni.

Del tempo vorreste fare un **fiume** per sostare presso la sua riva e
guardarlo fluire.

Ma l'eterno che è in voi sa che **la vita è senza tempo**

E sa che **l'oggi non è che il ricordo di ieri, e il domani il sogno
di oggi.**

E ciò che in voi è canto e contemplazione dimora quieto entro i confini
di quel primo attimo

in cui le stelle furono disseminate nello spazio.

Chi di voi non sente che la sua forza d'amore è sconfinata?

E chi non sente che questo autentico amore, benché sconfinato,

è racchiuso nel centro del proprio essere,

e non passa da pensiero d'amore a pensiero d'amore,

né da atto d'amore ad atto d'amore?

E non è forse **il tempo, così come l'amore, indiviso e immoto?**

Ma se col pensiero volete misurare il tempo in stagioni,

fate che ogni stagione racchiuda tutte le altre

**E che il presente abbracci il passato con il ricordo,
e il futuro con l'attesa.⁹**

A questo punto chiedo un'ultima concessione: convinto che ciascuno di voi avrà avuto modo di trovare nelle proprie conoscenze letterarie un percorso che completi la lacuna temporale di cui mi farò responsabile, saltando nel '900 e tralasciando possibili appelli a testi di età intermedia (in futuro il lavoro di tante persone potrà fornire ai materiali archiviati una continuità maggiore, speriamo), mi permetto di proporre alcune riflessioni non impertinenti su quello che si definisce 'Il sentimento del tempo'. In questo, occorre che ciascuno colmi coi mezzi propri della memoria individuale quanto si percepisce come 'eredità' del passato: solo così potrà apprezzarsi lo sforzo di mostrare il possibile punto d'arrivo della proposta didattica.

Lungo i decenni del secolo scorso, il tempo non è più quello quantitativo e lineare della storia, ma quello qualitativo, contorto e soggettivo della psicologia. Dalla realtà del tutto suddivisibile di Manzoni, Verga, Zola si passa a una sua frantumazione, che implica la frattura dell'io (*le moi partagé*), con la conseguente importanza attribuita alla memoria, capace di mantenere una (solo apparente) giuntura fra le parti, nel nome di una incrollabile fiducia dell'uomo in se stesso. Sciogliendo ogni legame tra causa ed effetto, cancellando il significato di spazio e tempo come lo aveva ricevuto nei secoli, l'uomo del '900 trovava in sé i parametri per cogliere il prima e il poi, anzi per annullarli negandoli, ribadendo invece, che solo la nostra memoria, il libero fluire della coscienza, consente di superare il senso di finitezza che ci angoscia e fa cogliere la morte in avvicinamento, attimo dopo attimo.

Quando una briciola di presente ci sollecita il ricordo, quando la memoria diventa 'vita' (o forse meglio, necessaria se non fondamentale e irrinunciabile illusione di essa), si annullano d'incanto allora le dimensioni temporali, viviamo un 'presente' che però tale non è, dal momento che vi è in esso la memoria, e ci serviamo di un 'passato' confuso nel 'presente'.

La 'memoria involontaria' è responsabile di questi attimi ove percepiamo nettamente la rottura della linearità temporale, l'insinuarsi del libero fluire della coscienza, capace di unire i cocci di una esistenza trascorsa, disseminati in un'area infinita, che sta tra il conscio e l'inconscio della nostra psiche.

Il punto di frattura tra un tempo circolare, ove principio e fine si identificano, e uno lineare, fatto di sequenze (il prima e il poi), sembra avvenire nei primi decenni dell'Ottocento, con Leopardi. Uno specialista come Bruno Biral sostiene che «i piccoli idilli introducono per la prima volta il senti-

9 Anche se non mancano gli influssi di Platone e Aristotele (per cui vedi *supra*), si nota la presenza qui di Eraclito (91 Diels-Kranz): «Non si può discendere due volte nel medesimo fiume e non si può toccare due volte una sostanza mortale nel medesimo stato, ma a causa della impetuosità e velocità del mutamento essa si disperde e si raccoglie, viene e va».

mento del tempo nella lirica italiana» (Biral 1965, 1158 e ss.). Non voglio discutere l'affermazione, che esclude poeti come Petrarca, Ciriaco de' Persi o altri secentisti, o lo stesso Foscolo: il sentimento del tempo è presente in modo più o meno marcato nella lirica di sempre; semmai Leopardi mostra di esserne animato senza soluzione di continuità, da quando ha scoperto il freddo vero, che ha segnato per lui il passaggio al nulla. Il tempo che scorre inesorabilmente fissa la conclusione di un fatto, che non può più ritornare, e si pone come sola realtà eterna, che porta alla percezione insopprimibile del vuoto, come si evince da *L'Infinito*:

E come il vento,
 odo stormir tra queste piante, io quello
 10 infinito silenzio a questa voce
 vo comparando: e mi sovvien l'eterno
 e le morte stagioni e la presente e viva
 e il suon di lei.

In questi versi è evidente il passaggio dallo spazio infinito al tempo eterno, che ci confonde e ci smarrisce; il tempo che procede 'involviendo' tutte le cose nel suo movimento comporta l'orrore al quale Leopardi reagisce delegando al ricordo la funzione di creare le illusioni, con lo scopo di superare la finitezza della dimensione temporale, nel mentre esse ribadiscono automaticamente, al subentrare del disincanto, l'impossibilità per l'uomo di sfuggire alla realtà, dunque l'obbligo di vivere sempre in compagnia della noia: un «desiderio puro della felicità, non soddisfatto dal piacere, e non offeso apertamente dal dispiacere» (*Dialogo di Torquato Tasso e del suo genio familiare*). In questa angosciante e angosciata esistenza l'uomo viene travolto dal fluire del tempo, per lui c'è un presente solo 'apparente', percepito e vissuto nel dolore, nella coscienza della precarietà a fronte dell'approssimarsi della morte, vivendo solo il passato, e anche questo in modo illusorio attraverso il ricordo. Ciò che 'viviamo', dunque, è un 'presente' fatto di passato, un passato che sa di fine, perché mai ci è dato di cogliere col ricordo un'esperienza nella sua totalità. E le cose, gli oggetti della realtà esterna, rimangono lì a rispecchiare il nostro essere nulla nel nulla. Così ancora Leopardi: «Io era spaventato nel trovarmi in mezzo al nulla, un nulla io medesimo. Io mi sentiva soffocare, considerando e sentendo che tutto è nulla, solido nulla» (*Zibaldone di pensieri*, 85). In questo percepire di essere nulla, cioè che la vita è morte, si accende, talvolta improvvisamente, confondendoci, anche, la scintilla del ricordo, che ci rinvia a ciò che siamo stati, facendoci definitivamente capire che viviamo esclusivamente del nostro tempo interiore. «Memorie della mia vita. Cangiando spesse volte il luogo della mia dimora, e fermandomi dove più dove meno o mesi o anni, m'avvidi che io non mi trovavo mai contento, mai nel mio centro, mai naturalizzato in luogo alcuno, comunque per altro

ottimo, fintantoché io non aveva delle rimembranze da attaccare a quel tal luogo, alle stanze dove io dimorava, alle vie...» (*Zibaldone*, 23 luglio 1827; il rinvio a Seneca per la prima parte è d'obbligo).

Una consapevolezza leopardiana, questa del poter vivere solo il nostro tempo interiore, che pare anticipare Proust; ma la stessa 'durata' di Bergson sembra già presente nel poeta di Recanati, quando afferma che le cose possono acquistare un loro valore, pur senza cavarsi di dosso il senso di nullità, quando su di loro si deposita e forma quasi una crosta la serie dei ricordi, che sia pur per un attimo riescono ad animarle. Quest'ultima considerazione restituisce al pensiero di Leopardi un atteggiamento affatto rinunciatario nei confronti del tempo, tale da fargli cogliere l'intima essenza e viverlo come e quanto ci è dato, senza deporre mai la consapevolezza che con esso e in esso viviamo la nostra morte.

Da Leopardi a Montale. «Il Montale è un poeta che non sa fermare il tempo» (così Biral 1965, 1166): nel fluire del tempo egli coglie la propria consunzione, l'insopprimibile sensazione che un continuo logorio avvulpi le cose in blocco. Neppure il ricordo può sottrarsi, e quando recupera dal profondo una parte del nostro vissuto, esso appare diverso, carente, confuso rispetto a ciò che è stato. Significativo, in questo senso, il messaggio di *Cigola la carrucola nel pozzo*, dove il poeta usa la similitudine tra il recupero della memoria e l'acqua (il tempo) che un secchio attinge dal pozzo (i ricordi), parte della quale, agitata dalla caduta del secchio, fuoriesce prima di giungere in superficie. Ma i timori che anche il ricordo 'muoia', lasciando in noi la sola certezza del nulla, sono tutti espressi in quella che viene vista come la poesia più significativa sul tema:

Non recidere, forbice, quel volto,
solo nella memoria che si sfolla,
non far del grande suo viso in ascolto
la mia nebbia di sempre.

- 5 Un freddo cala... Duro il colpo svetta.
 E l'acacia ferita da sé scrolla
 il guscio di cicala
 nella prima belletta di Novembre.

Si noti che il verbo iniziale indica già la separazione, segna la violenta cesura, simboleggia la dissociazione dell'uomo, temuta come un'esecuzione (cf. Ramat 1965, 115): tagliare il sottile filo che lega al ricordo è 'uccidere' l'illusione che ci anima, come Leopardi insegna. La forbice del tempo non deve tagliare quel volto, unico segno rimasto «nella memoria che si sfolla», che potrebbe confondersi e diventare la «nebbia di sempre».

Come abbiamo sentito, Eraclito diceva che non è possibile bagnarsi due volte nella stessa acqua del fiume; Montale sa che nel fiume del tempo non

è possibile bagnarsi nemmeno una volta, poi che il presente è illusorio, quanto il passato e il futuro.

La lotta tra il ricordo aggredito dal fluire del tempo, il sentimento del tempo stesso, sono i temi più evidenti del pezzo dal titolo *La casa dei doganieri*.

Tu non ricordi la casa dei doganieri
sul rialzo a strapiombo sulla scogliera:
desolata t'attende dalla sera
in cui v'entrò lo sciame dei tuoi pensieri

5 e vi sostò irrequieto.
Libeccio sferza da anni le vecchie mura
e il suono del tuo riso non è più lieto:
la bussola va impazzita all'avventura

e il calcolo dei dadi più non torna.
10 Tu non ricordi; altro tempo frastorna
la tua memoria; un filo s'addipana.
Ne tengo ancora un capo; ma s'allontana

la casa e in cima al tetto la banderuola
affumicata gira senza pietà.
15 Ne tengo un capo; ma tu resti sola
né qui respiri nell'oscurità.

Oh l'orizzonte in fuga, dove s'accende
rara la luce della petroliera!
Il varco è qui? (Ripullula il frangente
20 ancora sulla balza che scoscende...).

Tu non ricordi la casa di questa
mia sera. Ed io non so chi va e chi resta.

Composta nel 1930, questa poesia fa parte della raccolta *Occasioni e altro*; è dedicata a una giovanissima Arletta (o Annetta), protagonista di una breve ma intensa storia d'amore col poeta, e ora è scomparsa.

Affidandosi a un modo di esprimersi tipico, l'uso del linguaggio eliotiano correlativo-oggettivo, Montale ci dà per simboli il suo intendere il Tempo, la lotta per vincerne l'aggressione subdola, la lenta corrosione di tutto ciò che rimane all'uomo per credersi «immortale»; del ricordo fonte di illusione, via di fuga solo apparente e momentanea da una penosa condizione di sudditanza.

Gli «oggetti» creano i punti di riferimento più sicuri per darci la suprema illusione di riuscire a recuperare ciò che sembrava passato nel momento

stesso in cui era presente; ma quando anche questi perdono il loro aspetto consueto, allora la lotta diventa più dura, la bontà della situazione prima vissuta si trasforma in deludente, traumatizzante avversità.

La casa dei doganieri è l'emblema di tutto ciò. Posta in una zona di confine, che simboleggia il limite tra presente e passato, vita e morte, vita vera e non-vita, essa è soggetta alla violenza del Libeccio, dal Tempo che l'assale e l'allontana, distruggendone a poco a poco il ricordo. Simbolo di vita vera, quella concretizzata nel ricordo, la casa è ora «desolata», abbandonata nella realtà, «addolorata» nel sentire del poeta, e attende di far echeggiare ancora il piacevole suono della risata femminile, ma invano: il tempo ha vinto, la memoria non ha potuto conservare inalterata la felicità di quel momento. Subentra un profondo disorientamento, un frastornamento che dà la certezza, amara e dolente, di non poter giocare d'azzardo contro il caso, perché tutto si dissolve, lentamente e irrecuperabilmente, sotto l'azione irrefrenabile del tempo. Non resta che la netta sensazione della perdita, della definitiva separazione, della sconfitta, già comunque anticipata in testa alla poesia: il «tu non ricordi» è constatazione della negazione dell'essere, dell'esistenza stessa, perché togliere anche il ricordo significa togliere l'unica fonte di vita 'vera', che nasce dall'illusione. Altre volte il poeta ha chiesto che non si interrompesse il filo della memoria (*Non recidere forbice*), che lo lega alla vita, altre volte l'ha difeso, ha lottato per mantenerlo integro, ma... Qui la lotta è impari, il poeta non ha con sé la donna che lo possa aiutare a preservare il ricordo dall'assalto del tempo: esso è, purtroppo, unilaterale, e a nulla sembra valere la sua caparbieta nel tenere un capo del filo. In tanta negazione emerge, quasi con stupore, una 'luce' di speranza, un simbolo del 'varco', dell'uscita di sicurezza; un attimo, un flash che illumina a intervalli. La 'normalità' è ben diversa: è ancora e sempre il trionfo del non-essere, della non-vita, del non sapere. Ed in questo si fissa la condizione ultima dell'uomo di fronte al proprio sentimento del tempo: ancora disorientamento, tanto che non si riesce più a distinguere la vita dal suo contrario, tanto che ci si convince che il nostro vivere è un vivere la morte. Da qui l'eterno invito: *carpe diem*, anche se quel *dies* dovesse durare quattro milionesimi di secondo.

L'uomo da sempre guarda in se stesso, cerca di capire la sua intima essenza e in questa indagine si trova regolarmente dinanzi il proprio essere tempo: è questo il momento in cui egli percepisce la propria finitezza, ma, come gli è connaturato, non si arrende e mette in atto le tecniche più varie per cercare il tempo perduto. Marcel Proust vi dedicò un ciclo di sette romanzi, definito da Vladimir Nabokov «una caccia al tesoro, dove il tesoro è il tempo e il nascondiglio il passato».¹⁰ Anche attraverso le teorie bergsoniane (apprese direttamente dalle lezioni universitarie cui poté assistere),

10 Il giudizio fu espresso in una lezione su *Du côté de chez Swann*.

Proust mette in atto una «strategia attraverso la quale è possibile ridare vita al tempo che sembra irrimediabilmente 'perduto' nel meandro del passato: non la memoria volontaria e cosciente, controllata dall'intelligenza e dalla ragione, ma la memoria involontaria, scaturita da sensazioni inattese, da occasioni apparentemente insignificanti, che riesce a far emergere dal 'paese tenebroso' della psiche l'«immenso edificio» del ricordo, attraverso l'«evocazione di echi interiori e di associazioni intuitive».¹¹

La memoria poetica, in quanto 'ricordo', opera per cogliere la realtà, diventa strumento utile a creare un attimo di realtà, così come la poesia della memoria: e ci sentiamo 'in-finiti', proprio mentre cogliamo la nostra finitezza. Anche questo è un piccolo dono della ricerca intertestuale: la percezione che il tempo vinca tutto, ma pure la rasserenante certezza che - come sapevano gli antichi - esso non riesce a vincere la poesia, l'amore e la bellezza artistica.

Bibliografia

- Armellini, Guido; Colombo, Adriano (2000). *La letteratura italiana*. Bologna: Zanichelli.
- Biral, Bruno (1965). «Il sentimento del tempo. Leopardi, Baudelaire, Montale». *Il Ponte*, 21(8/9), 1156-76.
- Cresci, Flora (2000). *Percorsi per la terza prova (Scriptorium classicum)*. Firenze: Le Monnier.
- Fracassetti, Giuseppe (1869). *Lettere senili di Francesco Petrarca, volgarizzate e dichiarate con note*. Firenze: Le Monnier.
- Livio, Mario (2005). *L'equazione impossibile. Come un genio della matematica ha scoperto il linguaggio della simmetria*. Milano: Rizzoli.
- Paoli, Ugo Enrico (1969). *Orazio. I Carmi*. Firenze: Le Monnier.
- Peek, Werner (1955). *Griechische Vers-Inschriften*. Berlin: Akademie-Verlag.
- Pontani, Filippo Maria (1978-1981). *Antologia Palatina*, voll. 1-4. Torino: Einaudi.
- Ramat, Silvio (1965). *Montale*. Firenze: Vallecchi.
- Scarpa, Giancarlo (2000). *Nuovo Esame di Stato. Prima prova scritta. Esempi di svolgimento guidato*. Milano: Ciranna & Ferrara.
- Scarpa, Giancarlo (2014). «Per una classificazione tematica dei testi epigrafici e epigrammatici». Pistellato, Antonio (a cura di), *Memoria poetica e poesia della memoria. La versificazione epigrafica dall'antichità all'umanesimo*. Venezia: Edizioni Ca' Foscari, 11-26. DOI 10.14277/97735-95-3/SABP-3-2.

¹¹ Armellini, Colombo 2000, 7: 132; e per alcune delle osservazioni riportate sopra, anche Scarpa 2000.

Perceptual, Grammatical and Hermeneutical Dimensions of Digitality

Giorgio Buccellati

(Cotsen Institute of Archaeology, Los Angeles, USA)

Abstract Taking the work on the graphemic and morphemic analysis of the cuneiform texts of Ebla as a starting point, the paper reviews the 'grammatical' criteria that make digital coding not only more efficient and dynamic, but also intellectually more in tune with the goal of establishing an argument and unfolding a narrative. This throws light on aspects of software application on the one hand (such as the semantic web) and of the digital humanities on the other, ranging from textual to archaeological data.

Summary 1 Foundations. – 2 Perception. – 3 Grammar. – 4 Hermeneutics. – 5 Bibliographical note.

Keywords Digitality. Digital thought. Perception. Grammar. Hermeneutics. Prehistory. Semantic web. Narrative.

1 Foundations

If we consider seriously the prehistory of digitality, we are lead much farther back in time than one would expect. Such a perspective is helpful not because of an antiquarian interest, but because it helps us to become aware of unsuspected roots that deeply influence our stance today.

It is in fact far from irrelevant to go back to even the beginning of our species, over two and a half million years ago. A major feature that distinguished the first hominini from other members of the animal family was their ability to create contiguity where there was none. Shaping a stone into a tool entailed awareness of a structural template: on the one hand, the perception of the rough stone was accompanied by what, on the other, may be called a para-perception of the tool. The two perceptions, the stone and the tool, were present in the mind of the hominini before and during the execution. The inherent dynamics of this bracing of discontinuities had immeasurable consequences: it was the essential precondition of experimentation, which started our species on a road to progress that was unknown to other species.

The beginning of language signalled the next epochal step, some sixty thousand years ago. The bracing of discontinuities was extended to a

point where perception as such could be objectified in the form of words: one could conceive and name both the stone and the tool, irrespective of whether they were present to the speaker and/or the listener. Even more importantly, humans were now able to objectify, through syntax, the very bracing that linked elements that were otherwise distant from each other: they could linguistically describe the process itself of shaping the stone into a tool. Discontinuity extended, in other words, to a level where perceptions were no longer immediately tied to sensation. Perceptions and their dynamic correlation had acquired a conceptual and linguistic identity of their own: they were embodied in a reality, language, that was independent of the original reality to which it referred.

This new reality was, however, still tied to the physical confrontation between speaker and listener. Writing broke the barrier, some six thousand years ago, by providing what emerged as the extra-somatic embodiment of perception and of its linguistic conceptualization. The bracing of discontinuity was now placed outside the brain, in the physical medium that embodied the pertinent words. As a result, the potential for manipulation was incalculable. Thought had become fully reified and could be confronted as a thing in its own right, however closely linked to its original referent. A new contiguity of the discontinuous was now possible: a writer could place and a reader could see – physically side by side in their new graphic incarnation – things and processes that could not otherwise be seen.

In the written medium, such new contiguity was permanent, so much so that we can 'read' it millennia after it was written down. But it was static: it had to be activated by a reader. And here we can see how radical is the innovation brought about by digitality. While, at first, para-perception made the association of discontinuous perceptions possible; while, then, language and logic introduced the structuring of these associations on a somatic level; while, finally, writing provided a static extra-somatic extension of the same – digitality has now given us an *active* extra-somatic extension: structured perceptions can now interact with each other outside of the human brain. It is the only cultural innovation that can be placed on the same level as the other three epochal revolutions – para-perception, language, writing. The digital revolution is, indeed, an axial moment in cultural evolution.

There is a common thread that runs through the phases of development I have just outlined: it is the progressively greater distance among perceptions, on the one hand, and the equally progressive strengthening of the bracing power over them, on the other. At each step, both distance and bracing were magnified by being objectified through words and syntax first, then through the written signs, and now through digital programming. It is not as though perception could be bypassed at any one of these stages. Perception remained the controlling mechanism at all stages, starting the bracing and then judging the outcome. The finished stone tool, the

word and concept that expressed it, the written symbol that gave it a parallel tangible existence: they were all the result of a bracing of perceptions and they all, at the same time, offered sensations of a wholly new order for perception to internalise in turn. It is no different with digitality – as I will endeavour to show.

My personal research relating to the digital medium has been in the area of both textual and archaeological data from ancient Syro-Mesopotamia. With regard to the texts, I dealt with graphemic and morphemic analysis of the text from Ebla (dating to the 24th century BC), and a large corpus of Old Babylonian letters (18th and 17th century BC). I started at the dawn of the digital age, in 1968, and have continued throughout the remainder of what is by now almost half a century, establishing various grammars and writing pertinent programmes. With regard to archaeology, I brought the first personal computer to the field on an excavation in Syria in 1978, and I have continued to the present day through all the intervening stages of hard- and software development. I will not present here the technical aspects of these projects, which I did elsewhere. Instead, I wish to describe two factors that I found to be recurrent in my effort. I was drawn to reflect on these constants precisely because of the unrelenting process of change that was taking place in digital technology. In the fascinating whirlwind of innovation, which we eagerly adopted while just as eagerly waiting for what would next take its place, these two anchors remained firm, never overtaken. The first was the sense that human perception retained the commanding role at all times: the systems could not be auto-referential to the point of making human judgment irrelevant. The second was the realisation that the logic of structuring was upstream of all implementation, a logic that I describe under the heading of grammar. It is on the basis of these considerations that I will then draw some conclusions about the wider reach of digitality in terms of hermeneutics, especially in relationship to the humanities.

2 Perception

We will first address the question of perception. How does digitality impact the way in which we perceive the data? At its core, my answer relies on the historical (and *pre*-historical) argument developed above, concerning the bracing of discontinuities. Digital analysis confronts levels of discontinuity that are immeasurably greater and develops bracing mechanisms that are immeasurably more powerful than one could ever have expected before. And yet, the challenge in facing perception has remained the same. At the end of the line, processed data are offered to human perception and judgment, in ways that ultimately summon a confrontation similar to the one faced by humans and hominini in the other situations. The major

substantive difference is that the bracing itself is now inscribed in mechanisms that are outside of the human mind. Still, they call for a perceptual response that relies on an analogous human potential. We must assess more closely the nature of this confrontation.

Think of the size of the data we control, and of the degree to which we can articulate this control. In the case of Ebla, we have a total of some 173,000 words and almost 350,000 cuneiform signs; in the case of the Urkesh archaeological material, a total of almost 16 million records from only seven excavation units at the site. Each entry is tagged according to a rich categorization system, so that an equivalent count by attributes would reach a total many times over. Now, two observations help to put this in perspective, especially as we compare the perception resulting from digital processing with the one that immediately preceded it, based on physical card files.

1. In the digital world, we gain a perception of the whole because we can give precise totals that can be updated instantly; because we can just as instantly call for subdivisions and subgroups, adjustable with even the slightest change of parameters; and because we can see these groupings represented graphically in the form of pie charts or bar histograms. These results, and more, summon a specific perceptual response. The author caters to such perceptual dimension by seeking the best way to convey the nature of the corpus. The users expect this and adjust their response accordingly: they can criticise the formulation chosen by the author because such a reorganisation of the material is by now second nature to all. Our perception of the very role of quantification has changed. The perceptual response that writing had introduced into the world of orality helps us understand this: thousands of animals listed on a cuneiform ledger gave a perception of a 'herd' that was independent of any perception resulting from a physical confrontation. The king who had never even been to where the animals were would know the amounts from the graphic equivalents given on a cuneiform tablet, through the further filter of a scribe who would interpret the writing. The physical appearance of the signs incised on clay provided a guarantee of its correspondence to the animals as physical referents: the perception of these signs had replaced the perception of the original. Farther back in time, when first introduced, language had made it possible to have words for individual animals and the ability to combine them syntactically into new logical wholes. Naming animals and describing the process of herding created a new perception of the world: it was as if animals and their clustering had been invented for the first time. Sheep and herds had existed all along: but 'sheep' and 'herd', as words and concepts, placed a

completely new configuration on the way in which the same reality was perceived and reckoned with.

2. On the other hand, we lose, in the digital record, any perceptual connection with the single digital elements seen in their totality. It is wholly beyond perception to see the millions of records of the Urkesh database, or the tens of thousand of the Ebla digital archive, in their juxtaposition, because they are deeply hidden in the vast construct of a multi-tiered website. You can of course summon any single element at will, and you can see them clustered in new and unexpected ways. But the notion that they are there is taken on faith, in terms of perception. It is never like looking at a large book, or a multi-volume series, and see even just from the spines of these volumes how sizeable the whole really is. There is, truly, no overall perception of a digital whole. This may be contrasted with the pre-digital eras. A ledger, whether on a large clay tablet or in the form of a printed spreadsheet, provides such an overall view of the whole. The individual daily records may be detailed in separate documents, but the ledger as such, even one with many pages and subsections, has a perceptual consistency of its own. In the digital dimension, elements that are not contiguous are brought together in such a way that a new contiguity emerges, which transcends the physical dimension. It is a contiguity that is recreated conceptually and offered up to a higher level of perception, one on which a new judgment is exercised. What holds the pieces tightly together, in their enormous quantity and differentiation, is the grammatical scaffolding that supports them.

3 Grammar

The underlying factor that is instrumental in bringing about this new contiguity is the enhanced ability to brace the distance. This does not happen in a vacuum. The linking of discontinuity is far from haphazard: rather, what is being brought together is predisposed to such bracing by virtue of its very nature. We may think of such a predisposition as a tensional factor that is inherent in the structural make-up of each element: the template of a stone tool is inherent in the rough stone, since the toolmaker has to assess the size, shape, hardness, etc., of the material at hand in order to derive from it a given finished tool; as a digital equivalent, the clustering of a number of objects into a class of, say, clay conical cups depends on the formal attributes applied by the potter and exhibited by the individual items.

There are, we may say, two parallel notions of grammar. The first is constitutive: it is the actual coherence of the data. The second is reflective or descriptive: it identifies and conceptually defines such structural

coherence. A conical cup tends of its own accord, at the time of its making, to be clustered with other conical cups: this is the constitutive grammar. Conversely, we conceptualise the attributes that are inherent in the data and, thus, establish the corresponding descriptive grammar. The term 'ontology' is quite appropriate in this respect: it refers to the intrinsic nature of the data as well as to the formalization we give of it. My choice of the term 'grammar' over 'ontology' is due to the fact that it more adequately emphasises the tensional dimension of the elements with each other (in addition, I prefer to restrict the term to its function in classical philosophy). To use an example from the grammar of English: the moment I start a sentence (e.g. 'When I first arrived in...'), a whole series of constraints is set in motion, which limit the possible ways of continuing the sentence. This is the tensional factor: the preposition 'in' tends to, or expects, a reference to a location; the temporal clause tends to, or expects, a given resolution, invoking some and excluding others. It is in the nature of all structures to have components that are so tensionally linked with each other – hence, the constitutive dimension of grammar –, whereas the descriptive dimension reflects the conceptualization we give of it.

A digital grammar raises the ability to identify the most minute filaments in this tensional network of relationships, and to fully control their dynamics. An essential component of this is the phenomenon of capillary nesting. As far back as the pre-linguistic stage, we can infer a taxonomic awareness that even then could become progressively more complex, i.e. more intrinsically nested. The various tool kits reflect internal subdivisions where the types are clustered depending on degrees of differentiation and of similarity at the same time. Language and then writing make it possible to relate to these categories in an increasingly more pronounced definiteness, whereby the nesting itself receives its own expressive identity. A blade will be nested with others depending on whether they are made of stone or metal, on the size and shape, on the function and style, etc.

What is immensely more developed in the digital system than ever before is the almost unlimited extent to which such capillary nesting can be implemented. This is a central dimension of grammar. Capillarity means that the tensional bracing can be seen in the form of an inverted tree, where progressively lower nodes subsume clusters of attributes that come closest to the original organisation of the data. Nesting refers to the way in which some nodes are subsumed under higher ones. The power of the digital approach is that through such a capillary system one can instantly reach the lowermost nodes and the single elements that are included therein, regardless of how complex the system is.

Capillarity ensures continuity within discontinuity. While discontinuity refers to the de facto separation among items, continuity refers to the presupposition of hidden tensional linkages among them. If the uppermost node is, let us say, the archaeological site, there is, at first blush, no con-

tinuity between it (the site conceived as a whole) and the smallest sherd found in a deep trench. Or rather: the continuity is presumed given that the sherd is part of the site, but there is no immediate and apparent way to trace it. That is what digital capillarity makes possible. Inscribed in a grammatical system that defines the tensional qualities of each element, it describes conceptually the equivalent of the nervous system that links together all the intervening stages from the topmost to the lowermost node. This is due to the tagging that a proper grammatical categorization system recognises in the data. An ideal morphing is not an arbitrary overlay, but rather the conceptual result of making reality transparent. The enormous flexibility of digitality makes it possible to increase such transparency to the utmost. We can highlight here four aspects of the process.

First, atomism. The minimal constituents must be really minimal. At the very start of the confrontation with the data, digitality reduces the urge to interpret: the single record exists in its most elementary quality, and only its immediate formal traits are registered. In this perspective, one can understand the significance for atomism of two diverse areas of interest. (1) In archaeology, the record emerges at its most disconnected from the excavation, and it must be kept rigorously in this condition to provide the best possible measure of objectivity. (2) The whole emphasis of the semantic web on individual records instead of documents builds on the same premise: the records are tesserae of a mosaic that has to be constructed subsequently, and not concurrently, with the definition of the single record.

Second, predictability. A grammatical system is predictable in the way in which the individual elements relate to each other. It is for this reason that the elements, starting from the atomistic nature of their original disposition, fall into a unity. Thus, the application of a proper grammar to the archaeological record of an excavation makes it possible to predict, and therefore to project, a variety of ways in which the elements, as atoms within the system, can cohere into a tiered system of meaningful wholes.

Third, exclusivity. Exclusivity is an important result of predictability. The relationship among elements is univocal, i.e. 'exclusive', in terms of what it tends to and what it excludes. The phrase 'I run...' excludes a large number of possible ways to complete the sentence (e.g. it cannot be followed by 'yesterday'). A fundamental heuristic benefit of this is the opportunity of making statements of impossibility as to the occurrence of given relationships. The ultimate validity of such statements hinges on the size and quality of the sample and, thus, must be calibrated accordingly.

Fourth, automation. Given predictability and exclusivity, i.e. given a proper definition of the categorization system, one can automatically make explicit for each item a given attribute cluster that is otherwise implicit. Just as the phrase 'I run' is immediately understood as having the attributes of first person present of a given verb, so a programme can extract

from a textual string or from the digital representation of an artifact a wide set of attributes that establish correlations with all other strings and artifacts. This requires a special grammatical attention to the quality of the input: it has to be conceived as embedded in the same structural system that the grammar defines and can then operate on

These aspects of a digital grammar explain why the search function is such a prominent feature of digitality. One can reach for the most minute individual items, and for their clustering into complex groupings, in answer to the most specific question we may want to pose. That is the hallmark of the use of the digital medium in all its forms. But there is another aspect that is potentially just as far-reaching, but is not sufficiently implemented, especially in the humanities and social sciences. It is the opportunity of having an argument develop from the way the fragments are reconstituted. In section 1 above, I have stressed the active dimension of digital data in comparison to a written non-digital record, and that is because digital data are processed outside the mind. But the data so organised remain passive in one respect, namely because the building of an argument is generally left to the querying and the sorting that one applies to the data, since there is no unfolding narrative in a database. The underpinning of the grammar makes this possible, to the extent that the data can be organised automatically along progressive lines of significance. It is in this respect as well that the theory of a semantic web operates: reconstituting the individual records as fragments in an ordered sequence that follows the logical order of a sequential narrative, thus proposing more than a series of reconfigured clusters of data.

4 Hermeneutics

Grammar is essentially inner-referential: the coherence of a system is described in function of its own structure. Extra-referentiality, on the other hand, means that the value of a system is in function of a referent that is external to the system itself. Of this, there are two distinct types.

In the first type, the referent is chosen by the observer. It is fully defined from the start, which means that the nature of the referential range is under the control of the observer. To give a banal example: if we want to assign a chromatic value to a given set of items, and we choose a given scale external to the set of data, e.g. a Munsell colour chart, then a judgment about the interpretation of the colours stands and falls only with reference to that specific scale. It is the approach that is normally described as *-etic* or, better, *(e)-tic*, given the etymology of the suffix.

In the second type, the referent is assumed to be given within the cultural system from which it originates. Thus, the structural system of the three colours red-yellow-green at a street intersection depends not on the

precise chromatic value of the colours, but on a convention that builds on the contrast among the colours. The external referent is in this case the control of the flow of traffic – something that is second nature to us as being part of the culture that has designed the system. For an extraneous observer, the judgment as to the validity of the referent would not be immediately transparent. It would have to be inferred as to its effect from various observable correlations (the traffic stops at red, etc.), and as to its deeper nature in the measure in which it can be deduced from higher level correlations, which are not necessarily observable (in the case of the traffic light, the deeper nature is a convention accepted by vehicular traffic and not, for example, the result of a hidden power of the traffic light device). The progressively greater distance from observable facts makes interpretation less and less assured: herein the hermeneutic risk.

The first type of extra-referentiality is, as a matter of fact, an expanded grammar, i.e. a grammar where the external referent becomes an element of the grammatical system to which it refers. The correlation is grammatical in nature because, by positing the referent, the observer institutes a set of correlations that is univocal and explicit. Thus, adopting a colour scale to refine the taxonomic organisation of a given class of items means that the grammar is expanded to include a new roster of variables and lexicon of variants.

The second type of extra-referentiality, on the other hand, is specifically hermeneutical. Hermeneutics proposes factors that are presumed to be the hidden spring for a given constellation of factors. The case of the traffic light is simple and obvious to us, but for an observer wholly outside our culture the notion of a convention would have to be identified as being the factor that explains the various aspects of the flow of traffic. Here lies the hermeneutic risk, that is intrinsically a part of the extra-referential system for which the external referent is not posited by the observer.

These considerations have important consequences for the central themes of our *Convegno*. We may discuss them from the point of view of digital humanities. The concept is generally understood in the sense that data of a certain type (especially texts, and generally ancient texts) are processed through a variety of programmes. One aims for greater complexity of the data and for more sophistication in the way in which they are processed. The progress in the last few decades has been astounding. What has expanded immensely beyond limits that were long since established is the grammatical approach to the data: through programming, we propose ever new, and ever more challenging, parameters that function as external referents in re-organising the data. Being grammatical, it proposes a mechanical certainty that is in line with the model of the 'exact' sciences.

It goes without saying that this is of primary significance, and I have myself engaged fully in the venture. But it has often obscured what is even more important, namely the hermeneutical dimension of digitality, to which

we may refer as digital humanism. What makes the effort humanistic is not the nature of the data (such as texts), but rather the approach that accepts the risk of an external referent that is not posited by the observer. The role of digitality is then to present to the observer the broadest range of correlations on which the power of judgment can thus be applied. It offers, as it were, multiple re-configurations of the data to the observer's perception, which remains the judge. Herein lies, in my view, the unsurpassable limit to artificial intelligence, and the profound uniqueness of a humanistic mode of thought. I have given the definition on 'digital thought' to this new perception that digitality makes possible: it is the mode of thought that relates to the immensely powerful means we have to master discontinuities and rearrange them into a kaleidoscope of new contiguities.

We may thus better appreciate my initial comments about the foundations of digitality, which are in effect the foundations of digital thought. The progress of our species has rested on the ability to brace discontinuities in nature. What digitality has done has been to bring such bracing to the nth power. Conversely, this also shows the continuity across the vast gulf of time between us and the forebears of our species. Risk and experimentation have rested all along on the ability to brace discontinuities and to draw inferences from that. We are still progressing on the same road.

The title of our meeting, *Risorse digitali e strumenti collaborativi per le scienze dell'Antichità*, offers an apt formulation of the points I have been raising. A scientific approach to antiquity requires a grammatical structuring of the data, and 'resources' and 'instruments' are mechanisms that offer up to perception the results that have been thus made possible. The logo of the *Convegno* is evocative of the same: a library that seems to progressively vanish in the distance, as if escaping the control of the digital tablet in the foreground, is shown. The exact sciences would know only grammar, only the tablet in the foreground; they would know no hermeneutics. Digital humanism starts from the tablet, from these immensely powerful means of control over the data. But it then looks in the depth, accepting the risk ensuing from a reduced measure of control, but a richer appropriation of perceptions.

Bibliography

Rather than focusing on the specifics of individual systems, I have chosen to give in this article the conceptual and methodological foundations of a long research I have developed over the years, which included my direct involvement with the writing of various dedicated suites of programmes. These specifics I have dealt with in the publications listed in the relevant websites to which I give a reference below.

Of the pertinent literature, I will cite here only the most recent book by Federica Frabetti, *Software Theory. A Cultural and Philosophical Study*, London-New York 2015: it gives a thoughtful assessment of current trends and develops an original approach to the concerns I address here, starting from however different presuppositions and data.

For my work on digital projects in general I refer to the extensive entry on digital analysis in my personal website (giorgiobuccellati.net).

For the digital analysis of graphemics and morphemics in textual data, in particular with regard to the Ebla archives, see the website cybernetica-mesopotamica.org.

For the digital treatment of archaeological data from the excavations at Urkesh see the website urkesh.org.

The, see: Buccellati, Giorgio (2017). *A Critique of Archaeological Reason*. Cambridge: Cambridge University Press. URL critique-of-AR.net (2017-11-13).

Parte 3.

Profili organizzativi e giuridici

La protezione dei dati negli archivi digitali

Ignazio Zangara

(Università degli Studi di Catania, Italia)

Abstract The author describes and examines three subject approach sets: the first, technical-juridical, is related to methods of data transformation and conservation to guaranty protection of authenticity and access over time, respecting the law. The second, material, is related to physical protection data in terms of integrity and duplication. The last, purely juridical, pertains to data protection originating by intellectual work in copyright topics *lato sensu*. The essay exhorts to make considerations about the sharp contrast between juridical system prescription and how data is managed on line. The interests in rights protection by the authors and the data philosophy openly available – as it already occurs – needs agreement also in juridical system.

Keywords Copyright. Copyleft and public domain. Digital archives. Database protection. Open data. Creative commons. Juridical protection of data.

Gli studi sul diritto d'autore si sono spesso concentrati sulla tutela delle pubblicazioni tradizionali nei formati cartaceo e digitale (libri, collane, articoli) e la letteratura in proposito è amplissima. Al contrario, la riflessione sulla protezione dei dati contenuti in archivi digitali è ancora frammentaria e, pertanto, meritevole di ulteriori approfondimenti. Proprio su questa ho condotto la mia analisi per cercare di gettare nuova luce su una tematica così importante, le cui sfumature segnano i confini della tutela giuridica.

Un chiarimento lessicale è utile per individuare l'ambito di questo intervento; in generale, per 'archivi digitali' si intendono sia gli archivi storici (archivi di Stato, archivi comunali, militari, ecclesiastici, ecc.) – frutto di una operazione di digitalizzazione dal formato analogico originario, più o meno ben resa, e per i quali esiste un'apposita normativa e precise regole e di cui mi occuperò solo marginalmente – sia le raccolte di testi, di immagini o di dati informativi in genere, metodicamente disposti al fine di consentirne il recupero e, in questo secondo caso, mi riferisco tanto alle basi di dati primarie – cioè quelle che contengono i full text o, in generale, i documenti ultimi – quanto alle secondarie – cioè quelle che contengono solo riferimenti a tali documenti (bibliografie, indici, citazioni, ecc.).

La 'protezione dei dati' negli archivi digitali implica almeno tre direzioni di analisi: una, tecnico-giuridica, relativa alle modalità di trasformazione e di conservazione dei dati per garantirne autenticità e accesso, nel rispetto delle norme ordinamentali; una seconda, del tutto materiale, relativa alla

salvaguardia fisica del dato in termini di integrità e di duplicazione; un'ultima, puramente giuridica, riguardante la protezione dell'opera intellettuale sotto il profilo del diritto d'autore.

La trattazione seguirà questa tripartizione – che riflette, a mio avviso in progressione verticale, la problematicità delle questioni affrontate – e procederà, in ordine, dalla prima elencata, che potremmo inquadrare nel concetto di trattamento dei dati informatici ai fini dell'accesso e della conservazione.¹

La gestione del dato digitale deve essere garantita nel continuo processo di evoluzione tecnologica. La fragilità dei supporti, l'evoluzione del software e la non convenzionalità sui formati costituiscono una minaccia per la persistenza nel tempo dell'informazione digitale, rendendola talora irriconoscibile, irreperibile o, semplicemente, trasformata. Con riferimento specifico ai dati pubblici, che sono spesso oggetto, anche se non esclusivo, degli archivi digitali in parola, la normativa nazionale è intervenuta più volte per porre rimedio ai problemi testé accennati e ha dettato alcuni precetti, tenendo conto sia della disciplina comunitaria sia delle raccomandazioni e dei suggerimenti del W3C.²

Nel 2005, il nostro legislatore ha riordinato buona parte delle disposizioni in tema di informatizzazione e, pur esprimendo principi per l'amministrazione pubblica, ha finito per dettare regole di carattere generale alle quali anche i privati hanno iniziato a conformarsi progressivamente. La codificazione in un unico provvedimento normativo – il Codice dell'amministrazione digitale (CAD), emanato con D.Lgs. nr. 82/2005 e modificato ed integrato ripetute volte –, fornisce, tra le altre, precise indicazioni in merito alla standardizzazione della struttura dei dati pubblici³ (utilizzo dei formati aperti e neutralità rispetto agli strumenti tecnologici), alla definizione delle procedure di condivisione degli stessi (*open data*),⁴ al riuso

1 L'articolo, scritto originariamente nel 2014, è qui riprodotto in versione lievemente rivisitata e arricchita per dare conto dei più evidenti aggiornamenti normativi e dell'effettivo andamento dei progetti citati.

2 In proposito, è possibile consultare sia la Direttiva 96/9/CE dell'Unione Europea sia la documentazione pubblicata dal W3C disponibile all'indirizzo <http://www.w3.org/TR/>.

3 Il processo di standardizzazione dei formati dei contenuti pubblici ebbe inizio allorché, in ossequio alla Racc. W3C del 10/2/1998, il nostro Paese introdusse specifiche norme secondarie – il DPCM del 31/10/2000, in tema di protocollo informatico e la Circolare AIPA n. 40 del 22 aprile 2002, in tema di redazione degli atti normativi – aventi lo scopo di uniformare sotto il profilo strutturale i provvedimenti amministrativi e legislativi con la marcatura in XML.

4 Si vedano in proposito gli artt. 50 e ss. del citato Codice dell'amministrazione digitale.

delle informazioni digitali⁵ e alla loro conservazione⁶ (artt. 40-44bis) nel breve, nel medio e nel lungo termine.

Con la modifica del dicembre 2012,⁷ è stato introdotto l'obbligo di effettuare valutazioni comparative per la scelta del software da acquistare, in favore di quello di tipo libero o a codice aperto,⁸ in una prospettiva di superamento dei limiti dei sistemi proprietari. Ulteriori disposizioni, più recenti, sono state inserite nel decreto legislativo nr. 33 del 2013,⁹ in cui sono elencate con estremo dettaglio le informazioni che devono essere pubblicate in formato aperto ai fini dell'interoperabilità e dell'accesso, elevando in tal modo a sistema di economia e di sviluppo la logica della condivisione e del riuso. Nel mese di giugno 2014, l'Agenzia per l'Italia digitale, sulla scorta delle disposizioni normative primarie, ha pubblicato per la prima volta le «Linee guida per la valorizzazione del patrimonio informativo pubblico (anno 2014)», orientando la PA verso la produzione nativa dei *linked open data*, su cui si tornerà in seguito.¹⁰

5 Per quanto concerne il riuso dei dati e dei programmi informatici, si vedano rispettivamente le disposizioni normative contenute nel CAD agli artt. 50-62-ter e 69-70. Ai dati pubblici sono assimilati in questo contesto anche quelli costruiti con fondi pubblici.

6 Per approfondimenti sulla gestione degli archivi digitali, si veda lo studio dettagliato di Guercio 2009, reperibile all'indirizzo [http://www.treccani.it/enciclopedia/archivi-digitali_\(XXI_Secolo\)/](http://www.treccani.it/enciclopedia/archivi-digitali_(XXI_Secolo)/). Le regole tecniche sulla gestione e sul processo di conservazione dei documenti informatici, ivi compresi gli standard, i formati, i metadati e le altre specifiche da osservare, si vedano le regole tecniche contenute nel DPCM del 3 dicembre 2013, emanate ai sensi dell'art. 71 del CAD.

7 Il D.L. del 18 ottobre 2012, n. 179, convertito con modificazioni dalla L. del 17 dicembre 2012, n. 221, ha disposto la modifica del co. 1 dell'art. 68 del Codice dell'amministrazione digitale.

8 Il medesimo concetto è ribadito anche nella più recente legge del 9 agosto 2013, n. 98, *Conversione, con modificazioni, del decreto-legge 21 giugno 2013, n. 69, Disposizioni urgenti per il rilancio dell'economia*, nota come 'decreto del fare', che al secondo co. dell'art. 13-bis recita: «Le pubbliche amministrazioni possono usare piattaforme e soluzioni di acquisto on line accreditate anche ponendole in competizione tra loro. Qualora vi siano prodotti *open source* che non comportino oneri di spesa, il ricorso ai medesimi prodotti deve essere ritenuto prioritario».

9 Il provvedimento in parola riguarda gli obblighi di pubblicità, trasparenza e diffusione di informazioni prodotte o messe a disposizione dalla pubblica amministrazione tramite i siti web istituzionali. È interessante notare come all'art. 48 del D.Lgs. in parola si demandi al Dipartimento della funzione pubblica la definizione dei «criteri, modelli e schemi standard per l'organizzazione, la codificazione e la rappresentazione dei documenti, delle informazioni e dei dati oggetto di pubblicazione obbligatoria». Di fatto, l'Autorità Nazionale Anticorruzione (ANAC) aveva già emanato la Delibera n. 50 del 2013 con le «Linee guida per l'aggiornamento del Programma triennale per la trasparenza e l'integrità 2014-2016», in cui specificava alcuni criteri, modelli e schemi standard nei formati aperti (ad esempio, .rtf per i testi, .csv per i fogli di calcolo e .pdf/A per tutti i tipi di documenti).

10 Per la qualità dei dati delle amministrazioni sono previste le indicazioni derivanti dello standard ISO/IEC 25012 «Data Quality Model». In tal senso, in allegato alla Determinazione

Senza scendere nel dettaglio, la ratio delle norme appena accennate mira alla circolazione, alla condivisione, al riuso e alla protezione nel tempo del patrimonio informativo pubblico, prevedendo specifiche disposizioni, appunto tecnico-giuridiche, alle quali gli operatori devono attenersi.

Pur avendo citato soltanto alcune delle disposizioni inerenti al tema in discussione, il quadro normativo è assai ricco e si sta raffinando progressivamente. A ciò, tuttavia, non segue una corretta e rapida applicazione delle disposizioni, a mio avviso, per un duplice ordine di motivi: da un lato, perché non c'è il tempo di adottare i provvedimenti necessari per dare attuazione ad un disposto normativo perché lo stesso muta in qualcos'altro, non sempre in linea con il precedente. Si pensi alle vicende legate alla conservazione sostitutiva con i persistenti dubbi interpretativi¹¹ oppure a quelle, altrettanto dubbie, inerenti al deposito degli atti nel processo civile telematico;¹² dall'altro lato, a causa della scarsa confidenza dei più con gli strumenti avanzati sia sotto il profilo tecnologico sia sotto quello puramente concettuale, che presuppongono entrambi specifiche conoscenze ibride di tipo informatico e giuridico. Manca, neanche a dirlo, un piano di formazione per le nuove figure professionali. Non si tratta, infatti, di competenze acquisibili sul campo o, come direbbero oltreoceano, 'on the job', sono bensì vere e proprie conoscenze di metodo, oltre che di merito. E ciò è tanto più vero se si pensi che all'interno delle PPAA, ad esempio, impera ancora oggi il software proprietario, che grava in maniera pesante sulla spesa pubblica, solo perché quello libero imporrebbe l'acquisizione di competenze informatiche ulteriori che l'operatore, a qualunque livello, non ha intenzione di acquisire spontaneamente; la cartina di tornasole di questo concetto è che nel nostro Paese, tra i peggiori in Europa, sono poche le esperienze, e spesso relative ad un arco temporale determinato, di *linked open data*, o semplicemente di *open data*, di dati pubblici, per il

commissariale AgID n. 95/2014 del 26-06-2014, sono state pubblicate le *Linee guida per la valorizzazione del patrimonio informativo pubblico (2014)*, reperibili al seguente indirizzo web: <https://goo.gl/FkZYQx>, quali strumenti di indirizzo per le PPAA verso un processo di produzione e rilascio dei dati standardizzato e interoperabile. Dette *Linee guida* sono il frutto di una precisa delega da parte del legislatore del CAD che, al co. 7 dell'art. 52, ha previsto la seguente disposizione: «L'Agenzia definisce e aggiorna annualmente le linee guida nazionali che individuano gli standard tecnici, compresa la determinazione delle ontologie dei servizi e dei dati, le procedure e le modalità di attuazione delle disposizioni del Capo V del presente Codice con l'obiettivo di rendere il processo omogeneo a livello nazionale, efficiente ed efficace. Le pubbliche amministrazioni di cui all'articolo 2, co. 2, del presente Codice si uniformano alle suddette linee guida».

11 Sul punto, si vedano Lisi e Ungaro 2014, al seguente URL: <https://goo.gl/MU4aaj>.

12 Sul punto, si veda Minazzi 2014 al seguente URL: <https://goo.gl/fdteMn>. E inoltre Scorza 2015 al seguente URL: <https://goo.gl/dStkoh>.

solo fatto che la pubblicazione degli stessi non assolve, agli occhi di molti,¹³ ad alcuna utilità immediata ed è distante dal concetto di appartenenza/proprietà al quale per ragioni culturali bieche si resta molto legati. Alcuni, ancora deboli, recenti provvedimenti normativi¹⁴ stanno incoraggiando le PPA a pubblicare dati aperti, dietro la spinta della trasparenza e con l'incentivo collegato alla performance.

Con la seconda prospettiva di analisi accennerò ad alcune riflessioni sugli accorgimenti tecnico-informatici adoperati per proteggere 'fisicamente' i dati pubblicati.

Con la seconda prospettiva di analisi accennerò ad alcune riflessioni sugli accorgimenti tecnico-informatici adoperati per proteggere 'fisicamente' i dati pubblicati.

Sono definiti *Digital Right Management* (DRM) quei sistemi tecnologici utilizzati per la gestione dei diritti sulle opere digitali, sotto il profilo, per così dire, 'logico'. Si tratta, più segnatamente, di accorgimenti tecnici che i produttori possono inserire nelle opere digitali, o nei dispositivi atti alla riproduzione di esse,¹⁵ quali metadati e sistemi crittografici, al fine di condizionare il lancio o la riproduzione di file protetti al verificarsi di determinate circostanze. Alcuni espedienti di marchiatura dei contenuti digitali sono, ad esempio, il *product key*, avente lo scopo di sbloccare l'esecutivo del programma di riproduzione, o il *bitstream watermarking*, volto a limitare ad un numero predeterminato di volte l'esecuzione di file protetti e per limitare il numero di dispositivi sui quali i file protetti possono essere riprodotti, o il *fingerprinting*, per identificare l'utente finale del prodotto. Dello stesso segno sono anche le limitazioni delle funzioni di riproduzione, di copia/incolla, di stampa, di modifica di un documento in formato PDF, predisposte, ad esempio, dal noto programma *Acrobat* della *Adobe*. Tutti i sistemi di DRM sono deboli palliativi che hanno stuzzicato le abilità di migliaia di informatici sparsi nel pianeta al punto che oggi alcuni software gratuiti¹⁶ online trasformano in pochi istanti documenti protetti in formato

13 In proposito, si vedano Andreoli-Versbacha, Mueller-Langer 2014, 1621-33. E, inoltre, Donovan, Osborne, Watson 2015, 4-25.

14 Si vedano, tra gli altri, in proposito, il D. Lgs. nr. 150/2009; il D.Lgs. nr. 235/2010; il D.L. 201/2011; il D.L. nr. 83/2012; il D.L. nr. 179/2012; il D.Lgs. nr. 33/2013; la L. nr. 128/2013; il D.L. nr. 102/2014; il D. Lgs. nr. 102/2015; il D. Lgs. nr. 179/2016.

15 L'art. 102-quater della legge sul diritto d'autore autorizza espressamente l'apposizione di misure tecnologiche sulle opere digitali, destinata ad impedire o limitare atti non autorizzati dai titolari dei diritti.

16 Al seguente indirizzo <http://www.pdfunlock.com/it>, con una semplice procedura, è possibile sbloccare qualunque restrizione in un file PDF. In generale, sono abbastanza semplici le procedure da seguire per eliminare i DRM dai contenuti digitali, come è spiegato, ad esempio, al seguente indirizzo <http://www.semanticstone.net/il-digital-rights-management-drm-e-gli-ebook/>.

PDF in documenti totalmente aperti, editabili, riproducibili, stampabili, ecc. E con i nuovi orientamenti della Corte di Giustizia Europea¹⁷ non bisogna neanche mascherarsi da hacker per aprire i 'lucchetti elettronici' delle console o utilizzare *modchips* per attivare funzioni aggiuntive o aggirare protezioni sui dispositivi.

La maggior parte delle banche dati prevede dispositivi di selezione dell'utenza all'ingresso (user id e password), ma oltrepassata tale barriera la totalità dei dati sono fruibili indistintamente. Conosco pochi casi in cui vengono separate le categorie di utenti e diversificati i permessi di accesso ai dati.¹⁸ Tecniche, queste ultime, che consentono di proteggere in maniera selettiva i dati pubblicati in considerazione della tipologia: dato pubblico e dato protetto.

Anche Google, che ha enormi risorse economiche e informatiche, per il progetto 'Google Books', pur aprendo al mondo intero una vasta porzione di produzione libraria, per mantenersi nei ranghi delle norme a protezione del diritto d'autore, ha dovuto limitare a monte la riproduzione di tutte le pagine dei testi, lasciandone disponibili soltanto alcune, proprio perché è ben noto che, in ambito informatico, la visualizzazione di un oggetto è sinonimo di riproduzione all'infinito dello stesso.

Porre limiti logici al flusso dei dati digitali nella rete è, come abbiamo visto, assai arduo e non produce i risultati attesi. Il sistema dei DRM, per sua natura rigido e penalizzante, è stato sottoposto a pesanti critiche;¹⁹ un limite *contra legem*, ad esempio, impedisce la libertà di utilizzo del file protetto anche al legittimo licenziatario, allorché questi intenda fare una copia di backup dell'opera, attività che, anche secondo la più ristretta normativa sul diritto d'autore, è assolutamente legittima e ha lo scopo di scongiurare la perdita dell'opera licenziata.²⁰ Tali tecniche di inibizione,

17 Si veda *infra* l'ultimo periodo della nota 20.

18 Senza scendere nel dettaglio, un caso in cui la categoria dell'utenza determina le modalità di accesso e la qualità dei dati forniti è il sistema *Emeroteca on-line* dell'Università di Catania, che raccoglie gli indici dei periodici delle discipline giuridico-economico-sociali di maggior rilievo. Altro esempio di selettività all'ingresso è *BIA-Net*, il cui accesso prevede quattro categorie di utenti con poteri di amministratore, di editore, di studioso e di ospite, allo scopo di migliorare la qualità e la quantità dei dati esistenti e di tutelare i diritti esistenti sul framework, senza impedire agli ospiti di visualizzare liberamente i dati pubblici. Nel caso in cui gli autori di una pubblicazione volessero mantenere il copyright (la cui accezione è squisitamente economica) sulla stessa, potranno rivelarsi molto efficaci le policy di distribuzione che prevedono l'accesso libero ad una porzione limitata dell'opera e il pagamento di un corrispettivo per la licenza d'uso del prodotto nella versione completa.

19 Sulla pericolosità dell'autocrazia dei controlli connessi ai servizi DRM e sulle censure e incompatibilità con altre licenze, si veda Peruginelli 2007, disponibile al seguente indirizzo <http://www.ittig.cnr.it/Ricerca/Testi/peruginelli2007.pdf/>.

20 In effetti, l'art. 615-quater del codice penale, rubricato «Detenzione e diffusione abusiva di codici di accesso a sistemi informatici o telematici», dispone che «Chiunque, al

come accennato poco sopra, per quanto sofisticate, costituiscono terreno di conquista per gli informatici che, di volta in volta, hanno sempre trovato il sistema per aggirarle, diffondendo in rete le informazioni necessarie affinché tutti, con minime competenze, possano beneficiarne. Di qui, l'inadeguatezza e l'inopportunità di tali strumenti di natura restrittiva.

In un panorama di dati aperti, come quello che si sta delineando negli ultimi anni, le protezioni logiche, oltre ad avere una portata assai limitata, risultano essere del tutto anacronistiche.

L'ultima prospettiva di analisi sulla protezione dei dati, quella squisitamente giuridica, è maggiormente problematica. Dinnanzi al progressivo affermarsi di nuove tecnologie dell'informazione e della comunicazione, nonché ad un diverso sentire da parte degli autori con riferimento alla tutela della propria opera intellettuale, l'impianto normativo non ha saputo fornire risposte adeguate. Le leggi vigenti vorrebbero, senza riuscirci, mantenere vincoli e limitazioni ormai inadeguati, tutelando l'opera dell'ingegno in maniera pressoché indistinta sia che si tratti di un testo sia che si tratti di una fotografia o di una tela dipinta sia che si tratti di un software sia che si tratti di una elaborazione del pensiero espressa in forma non convenzionale (blog, singoli contributi e raccolte multimediali, siti web, banners, landing pages), senza tenere conto della volontà dei produttori dello sforzo creativo e intellettuale o, per meglio dire, senza lasciare loro i necessari spazi di disposizione tali da consentirgli forme di tutela flessibili.

Mi riferisco ovviamente alla legge sul diritto d'autore (l. nr. 633/1941 e successive modifiche ed integrazioni), che da almeno un ventennio è oggetto di un'accesa discussione a causa di una prassi che spesso si pone in netto contrasto con i precetti ivi contenuti.

Nello specifico, proverò a riassumere cosa prevede la normativa nazionale sul diritto d'autore nei confronti di coloro che hanno investito tempo, risorse intellettuali ed economiche per creare, o anche semplicemente per

fine di procurare a sé o ad altri un profitto o di arrecare ad altri un danno, abusivamente si procura, riproduce, diffonde, comunica o consegna codici, parole chiave o altri mezzi idonei all'accesso ad un sistema informatico o telematico, protetto da misure di sicurezza, o comunque fornisce indicazioni o istruzioni idonee al predetto scopo, è punito con la reclusione sino ad un anno e con la multa sino a euro 5.164». Dello stesso avviso è il legislatore della seconda parte del 1 co. dell'art. 171-bis della legge sul diritto d'autore per quel che riguarda i programmi per elaboratore. Dunque, anche l'attività di rimozione o di semplice predisposizione di indicazioni o istruzioni volte alla rimozione delle misure di protezione di un sistema informatico - è il caso dei DRM - costituisce reato (nella seconda ipotesi si configura perfino una fattispecie di reato di pericolo). Tuttavia, un orientamento di apertura è già presente in una interessante sentenza della quarta sezione della Corte di giustizia dell'Unione europea del 23 gennaio 2014 (causa C 355/12) che dispone, in ossequio al principio di proporzionalità, la non perseguibilità della condotta volta alla rimozione dei dispositivi di protezione DRM, attraverso tecniche di *cracking*, se ciò avviene per uno scopo lecito. In altri termini, la pronuncia è assai significativa perché autorizza la rimozione dei dispositivi di protezione dei sistemi informatici nell'ipotesi in cui i contenuti da visualizzare siano licenziati o non siano protetti dal diritto d'autore.

aggregare, conoscenza²¹ e l'hanno espresso in formato digitale, soffermandosi solo sulle disposizioni riguardanti le banche dati.

La legge distingue i singoli dati, come entità/oggetti a sé stanti, dalle collezioni di dati che per scelta, disposizione e organizzazione delle informazioni assumono carattere di originalità e, infine, i dati messi insieme in assenza di una particolare selezione (una raccolta di documenti), nel caso in cui l'autore - che nell'ipotesi di specie prende il nome di costituente - per la sua realizzazione abbia sostenuto un 'investimento rilevante'.

Nel primo caso, l'ordinamento giuridico assegna la protezione piena²² qualora il dato sia il frutto di un'elaborazione intellettuale originale - e non entra nel merito del grado di originalità stessa (cioè non specifica la misura del *quid novi* rispetto a un'altra opera esistente) - e attribuisce al dato, appunto, la detta protezione solo a condizione che non si tratti di mera copia (in caso contrario, rientreremmo nel concetto di plagio). Mi sembra utile distinguere, a livello concettuale, il plagio dalla contraffazione. Nel plagio la condotta repressa è l'usurpazione della paternità di un'opera; nella contraffazione la condotta repressa è l'appropriazione di un'opera altrui al fine di essere sfruttata economicamente senza il consenso dell'autore. Per dirla in termini tecnici, il plagio incide sul diritto morale d'autore, la contraffazione sul diritto patrimoniale d'autore. Le pene della reclusione e della multa associate a queste condotte illecite sono specificate agli artt. 171 e ss. della legge sul diritto d'autore.

Proseguendo, per espressa previsione del legislatore, i dati raccolti nelle banche dati (ai nostri fini non è importante distinguere i *reference* dai *source database*) rientrano nella tutela del diritto d'autore, lasciando impregiudicati i diritti d'autore sulle singole unità documentali.

Sono intese anche di dati le raccolte di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo.

Sono oggetto di tutela non solo quei particolari elementi di carattere creativo inerenti alla concreta raffigurazione della banca dati (schermate, tabelle, indicizzazioni) ma, anche, l'eventuale singolare selezione dei dati (metodi di accesso ai dati) che sono stati ammessi nella compilazione ad opera dell'autore.

21 Il diritto d'autore, com'è noto, scaturisce contestualmente alla creazione dell'opera e consta del diritto morale e del diritto patrimoniale d'autore. Con il primo - che è imprescrittibile, irrinunciabile e intrasferibile - si intende tutelare la paternità dell'opera con una serie di diritti che spettano solo al creatore. Con il secondo - che è trasferibile ed ha una durata prestabilita dalla legge (per conoscere quale durata è assegnata al copyright nei diversi Paesi è possibile consultare al seguente indirizzo <https://goo.gl/rer56W>) - l'ordinamento giuridico riconosce al titolare il diritto di utilizzazione esclusiva dell'opera, salvo cessione del medesimo che avviene, di norma, dietro corrispettivo.

22 In sostanza, la legge assegna all'autore dell'opera dell'ingegno il diritto di esclusiva, riservandogli l'esercizio delle attività di produzione, traduzione, adattamento, trasformazione, modificazione e assoluta libertà nella forma di distribuzione al pubblico.

Si tratta, più in dettaglio, di collezioni di dati che sono talmente originali (per la disposizione, per il trattamento dei dati, per la selezione che rispetto ad una massa hanno subito) da assumere la dignità di opere dell'ingegno e quindi di oggetto della proprietà intellettuale in senso pieno (art. 3, par. 1, della Direttiva 96/9/CE del Parlamento europeo).

Per le banche dati, a parte la selezione e la disposizione dei materiali, un cenno deve farsi anche sul software di organizzazione, di visualizzazione e di ricerca che su quei dati opera, che assume di per sé la dignità di opera dell'ingegno. Detto in altri termini, altro è la struttura dei dati altro è la modalità attraverso la quale essi vengono classificati, estratti dal database, aggregati, presentati e visualizzati dall'utente. L'esempio, a chiarificazione di quanto appena detto, è un testo in HTML, che sta alla base di ogni pagina web; a video appare una pagina di testo, talvolta, con immagini e suoni che è il frutto di un codice - il linguaggio di marcatura, per l'appunto - che ha tutt'altra struttura e codificazione e che viene eseguito da un software per essere mostrato all'utente in forma esteticamente gradevole, funzionale alla lettura, che pone in risalto le parti più importanti, che consente al lettore di attivare strumenti quali il copia/incolla, la stampa, la memorizzazione, ecc. In questa stessa logica, quindi, nello strato base di una banca dati sono presenti diversi livelli di dati, tra cui il software di gestione e il software di ricerca, che hanno algoritmi propri e che quindi vanno distinti dai documenti, anche in termini di protezione del diritto d'autore.

Il sistema *BIA-Net* sui diritti dell'antichità è un esempio di opera complessa in cui si integrano più archivi con strutture di dati differenti, gestiti da programmi trasversali che permettono l'organizzazione e il recupero delle informazioni (cf. Spampinato, Zangara 2013). I dati del *Thesaurus* (l'archivio che contiene il vocabolario in linguaggio controllato dei concetti per l'indicizzazione dei documenti di *BIA-Net*) sono modellati secondo lo schema descrittivo *Simple Knowledge Organization System* (SKOS), che potrà essere fruito in differenti 'viste' secondo quanto disposto dal sistema di aggregazione e di *front-end* prescelti. In *BIA-Net*, quindi, sul livello 'codificazione dati' lavora il sistema di ricerca che si interfaccia anche con gli altri archivi aventi differenti linguaggi di codificazione (*Dublin Core e Text Encoding Initiative*) per essere raccolti e presentati all'utente secondo un determinato schema di *front-end*. Tutti questi elementi, fusi insieme nell'unico prodotto *BIA-Net*, sono insieme, ma anche autonomamente, meritevoli di protezione giuridica.

Da ultimo, la legge riconosce una tutela affievolita a beneficio di colui che ha semplicemente aggregato o raccolto dati, producendo un'opera del tutto priva di carattere di originalità. Questi che, come accennato sopra, prende il nome di costitutore - persino la denominazione 'costitutore', e non 'autore', rientra in una logica di diritto attenuato - per il solo fatto che ha investito tempo e danaro nella realizzazione della stessa, gode

della protezione da parte dell'ordinamento sulla sua opera. Tale ipotesi si inquadra giuridicamente (ex art. 7 della Direttiva 96/9/CE del Parlamento europeo) all'interno del cosiddetto diritto sui generis. Si hanno, in questo caso, semplici aggregazioni di dati e l'interesse perseguito dall'ordinamento giuridico è quello di tutelare, quanto meno, lo sforzo economico 'rilevante' sostenuto dal costituente nell'intrapresa.²³

La creazione di tale istituto di tutela affievolita è legata alla facilità con cui, tramite gli strumenti elettronici, è possibile estrarre e reimpiegare i contenuti delle banche dati con costi minimi, o persino nulli, vanificando lo sforzo economico sostenuto per la sua prima pubblicazione. In virtù del diritto *sui generis* il costituente della banca dati ha il diritto di vietare le operazioni di estrazione ovvero reimpiego della totalità o di una parte della banca dati, qualora tali operazioni non rientrino nella normale gestione dell'archivio o gli arrechino un pregiudizio ingiustificato.

Le attività di accesso e di consultazione, le eventuali operazioni di riproduzione permanente della totalità o di parte sostanziale del contenuto su altro supporto sono soggette ad autorizzazione da parte del titolare del diritto di utilizzazione economica (tipicamente, l'editore che cede dietro pagamento la licenza d'uso del prodotto).

Le attività di accesso e di consultazione, le eventuali operazioni di riproduzione permanente della totalità o di parte sostanziale del contenuto su altro supporto sono soggette ad autorizzazione da parte del titolare del diritto di utilizzazione economica.

Il legittimo utilizzatore, cioè colui che ha acquistato la licenza d'uso, ha libertà di estrarre o reimpiegare parti non sostanziali del contenuto della banca dati per qualsiasi fine. Una pronuncia della Corte di Giustizia, tuttavia, ha chiarito, estendendolo, il concetto di estrazione e reimpiego di parti non sostanziali, comprendendo altresì il «trasferimento di parti non sostanziali che, per il loro carattere ripetuto sistematico, abbiano condotto alla ricostruzione di una parte sostanziale» del contenuto di una banca dati.²⁴ In altri termini, altro è raccogliere le informazioni, anche ripetutamente, da una banca dati per eseguire lavori unici o di routine, altro è prelevare ciclicamente informazioni più o meno complete al fine di riprodurre parte sostanziale della banca dati, recando un pregiudizio al titolare del diritto *sui generis*. Delle due attività descritte, solo la seconda, ovviamente, si pone in contrasto con l'ordinamento giuridico.

Infine, il co. 1 dell'art. 64-sexies della legge sul diritto d'autore elenca due ipotesi di libera utilizzazione di una banca di dati. In particolare, la prima prevede che non è possibile limitare l'accesso o la consultazione della banca dati quando tali attività abbiano esclusivamente finalità

23 Per approfondimenti si veda Dal Poggetto 2007; De Robbi 2002, e Peruginelli 2006.

24 Corte di Giustizia UE, sez. IV, sentenza n. C-304/07 del 2008-10-09.

didattiche o di ricerca scientifica, non svolta nell'ambito di un'impresa, purché si indichi la fonte e nei limiti di quanto giustificato dallo scopo non commerciale perseguito; la seconda, prevede che non è possibile limitare l'impiego della banca dati per fini di sicurezza pubblica o per effetto di una procedura amministrativa o giurisdizionale.

Vado a riassumere quanto già descritto con un esempio che mi sembra abbastanza chiarificatore. La 'nota a sentenza' è un derivato, se mi si passi il termine, di un atto di pubblico dominio, la decisione del giudice. Essendo la 'nota a sentenza' il frutto di un'elaborazione intellettuale, essa assume la dignità di opera dell'ingegno in quanto documento dotato di originalità rispetto alla matrice (la sentenza). Gli autori delle 'note a sentenza', dunque, possono vantare la piena tutela *ex lege* 633/1941. Rimanendo nello stesso esempio, diverso, ovviamente, è il caso del dato 'sentenza'. Tale unità documentale è pubblica, pertanto, su di essa, singolarmente, nessuno può vantare diritti o pretese. Ora, non al dato pubblico, ma alla collezione di dati pubblici, continuando nell'esempio, quindi, una collezione di 'sentenze' collazionate in una banca dati - ma potremmo anche spingerci oltre, considerando i soli riferimenti ad esse, come le citazioni, ad esempio: Tribunale di *sede*, sent. n. *tale*, pronunciata in una certa data - è concessa la tutela giuridica affievolita *ex lege* 633/41, a condizione che l'impegno finanziario sostenuto dal costituente per la realizzazione dell'opera sia da considerare 'rilevante'.

Proseguendo, con un altro esempio, nel caso in cui un archivio digitale contenga sia dati pubblici che dati protetti dal diritto d'autore (torno al caso di *BIA-Net*, dove le fonti antiche del diritto sono libero patrimonio dell'umanità ma le unità documentali - record - che le contengono sono arricchite da ulteriori informazioni, quali i codici di classificazione, le opere bibliografiche di riferimento collegate con le altre fonti, ecc; lo stesso vale per l'archivio di bibliografia), stante la genericità delle norme, occorre trovare, con soluzioni tecnologiche, il corretto equilibrio per assicurare a chiunque la piena fruibilità di ciò che è di pubblico dominio senza perciò disattendere le aspettative di coloro che hanno investito risorse intellettuali ed economiche per la realizzazione dell'opera (v. *infra*).

La tutela ordinamentale rivolta all'autore o al costituente, prevista nel provvedimento a cui abbiamo appena fatto cenno, finisce per stare stretta allo stesso soggetto che dovrebbe beneficiarne e, per questo motivo, hanno acquisito peso, soprattutto per i contenuti digitali, anche altre forme di tutela straordinarie. Queste, come vedremo, permettono all'autore dell'opera o a chi ne detiene i diritti di sfruttamento economico (cf. Asante, Fagnocchi, Longo 2014) di trarre beneficio in termini di notorietà e, anche se non nell'immediato, in termini economici.

Nell'ultimo ventennio, invero, con il proliferare delle informazioni in rete, sia di prima mano che riciclate, la legge sul diritto d'autore ha iniziato a mostrare i suoi limiti; la tendenza è quella di rendere pubblici ed

accessibili i dati in una sorta di condominio della conoscenza per favorire condivisione e sviluppo di quell'intelligenza collettiva che la scienza tanto apprezza.²⁵

Autori e fruitori della conoscenza hanno in comune un macro-scopo fondamentale che è proprio quello dell'incremento e della diffusione del sapere nell'ecosistema digitale.

Gli *Open data* (OD) sono espressione di questo pensiero e sono in assoluta contrapposizione con le logiche tradizionali di protezione dei dati.

La strategia dei dati aperti è prevista, come abbiamo accennato sopra, come modalità di gestione della cosa pubblica, quale elemento di trasparenza e di apprezzamento in termini di performance degli enti pubblici.²⁶ Dati singoli e *dataset* informativi, sono sempre più spesso condivisi, se non gratuitamente, con costi marginali di riproduzione²⁷ al fine di innescare circoli virtuosi di produzione e di efficienza e sono considerati elementi chiave per lo sviluppo economico (ciò che nel mondo anglosassone viene descritto con l'espressione *trickle down effect*).

Le *Linee guida* dell'Agenzia per l'Italia Digitale (AgID), sopra citate, spingono la PA italiana a produrre *linked open data* (LOD) fornendo precise indicazioni su come elevare il livello di qualità dei dati rilasciati per poter essere utilizzati al meglio sia da operatori umani che da operatori virtuali (agenti intelligenti).

Il beneficio più evidente nell'utilizzo dei LOD è poter sfruttare la semantica dei dati e le infrastrutture web per abilitare la scoperta e l'inferenza di nuove informazioni, con ricadute positive e dirette sia per gli utenti che per gli sviluppatori. Così, dati propri di un ente e dati cosiddetti *mushup*, ossia raccolti da altri *data provider*, combinati e lavorati tramite software

25 In tal senso, alcune riviste digitali romanistiche hanno intrapreso una linea pressoché comune di condivisione del sapere; mi riferisco, ad esempio, sia alla rivista *Archeologia e Calcolatori*, diretta da Paola Moscati (<http://soi.cnr.it/archcalc/index.htm>), sia a *La rivista di diritto romano*, diretta da Ferdinando Zuccotti (<http://www.ledonline.it/rivistadirittoromano/>), sia alla rivista *Diritto e storia*, diretta da Francesco Sini (<http://www.dirittoestoria.it/>), che pubblicano gli articoli a testo pieno in open access. Di approccio diverso è, invece, sia la rivista belga *L'Antiquité classique* (<http://www.antiquiteclassique.be>) sia la rivista *Ivris Antiqui Historia. An International Journal on Ancient Law*, diretta da Gianfranco Purpura (<http://www.libraweb.net/riviste.php?chiave=312&h=427&w=300>), che pubblicano online gratuitamente gli indici e rilasciano il testo pieno dietro il pagamento dell'abbonamento. Per approfondimenti sulle policy di distribuzione dei dati contenuti negli archivi in rete di fonti antiche (letterarie, epigrafiche e papirologiche) e nei repertori bibliografici anch'essi in rete si rinvia al portale *BIA-Net – Il portale delle risorse digitali per il diritto romano e per le scienze dell'antichità*, sotto la direzione scientifica di Nicola Palazzolo, disponibile online (<http://104.236.71.119/bianet/>).

26 Sugli sviluppi dell'*open data* in Italia si veda Dominici 2014, disponibile online (<https://goo.gl/yvRQa3>).

27 La tariffazione è regolata da precise disposizioni normative contenute all'art. 7 del D.Lgs. n. 36/2006.

appositi, permettono agli stakeholder (studiosi, soggetti pubblici e privati) di fornire informazioni e servizi del tutto innovativi, customizzati, integrati. La Direttiva *Public Sector Information* (PSI), appunto, sull'informazione del settore pubblico (Direttiva 2003/98/CE), recepita in Italia dal D.Lgs. n. 36/2006, così come modificato dalla L. n. 96/2010, ha sancito il libero riutilizzo di tutti i documenti pubblici²⁸ anche per fini commerciali, oltre che per scopi non commerciali, pur mantenendo come unico limite il rispetto della disciplina in materia di protezione dei dati personali.

Sono già apprezzabili le applicazioni che lavorano, aggregano e riordinano LOD pescandoli da siti differenti per la *business intelligence*, per le *Smart Cities* o per anticipare e cavalcare le tendenze di mercato.

La filosofia dei dati aperti nasce sulla scorta della buona esperienza della licenza di tipo *GNU's Not Unix - General Public Licence* (in acronimo ricorsivo, GNU-GPL) per il software libero,²⁹ che ha permesso agli sviluppatori e ai semplici utenti di disporre, appunto liberamente (che è un concetto più ampio rispetto alla gratuità), del codice sorgente così come auspicato dal suo fondatore Stallman, sin dai primi anni '80, per diffondere libertà e cooperazione attraverso il web. In sostanza, la licenza consente di utilizzare, copiare, modificare e ridistribuire il software, perpetuandone la natura libera, qualunque siano le modifiche e gli adattamenti effettuati. L'autore potrà avvalersi della tutela connessa alla licenza GNU-GPL nel caso in cui altri, utilizzandolo, omettano di citarlo oppure nel caso in cui dal software derivino altri applicativi e questi vengano immessi nel mercato con licenze diverse da quella originaria.

Una soluzione di compromesso tra i sostenitori del diritto d'autore, nella sua espressione più ampia, e i fautori della libera circolazione del pensiero e delle idee tramite il web è rappresentata dalle ormai note *Creative Commons Public Licence* (CCPL). Da una decina di anni, gli autori di pubblicazioni digitali hanno in tal modo la possibilità di conservare alcuni diritti sulle proprie opere, concedendo agli utenti della rete³⁰ la libertà di riproduzione e di utilizzo a determinate condizioni. Dette licenze sono denominate *Attribution, Non commercial, No derivatives* e *Share alike* e possono essere combinate tra loro fino a sei soluzioni possibili e utilizzate da tutti liberamente e gratuitamente. Lo sforzo che l'omonima (CC) orga-

28 Sono escluse dall'applicazione del decreto in parola soltanto le categorie di documenti elencati all'art. 3.

29 Costituisce ulteriore esempio di software libero il noto Apache 2.0. Per maggiori approfondimenti, si veda la pagina online dedicata alla licenza al seguente indirizzo <https://www.apache.org/licenses/LICENSE-2.0>. A proposito delle distinzioni tra software libero, copyleft e software commerciale, si rinvia al seguente link <https://www.gnu.org/philosophy/free-sw.it.html>.

30 Le CCPL possono essere associate anche ai contenuti off line, anche se i supporti ottici e le memorie di massa esterne sono sempre meno utilizzate per contenere pubblicazioni.

nizzazione no profit statunitense, leader del movimento *copyleft*, insieme ai paesi affiliati al network, ha compiuto negli ultimi anni è stato quello di armonizzare la disciplina per ottenere licenze che siano compatibili con tutte le legislazioni locali e lo scorso novembre è stata pubblicata la versione 4.0, con *Commons Deed*, *Legal Code* e *Metadata*³¹ validi in tutti i pesi, senza necessità di specifici adattamenti di *porting*.³² Si tratta di uno strumento di protezione di alcuni diritti d'autore alquanto flessibile. Ad esempio, la licenza CCPL scelta per la distribuzione di un'opera non è esclusiva e all'autore è consentito optare per una diversa combinazione di CCPL per altro tipo di distribuzione della stessa opera. Ulteriore caratteristica delle CCPL è quella di essere 'human and machine readable'; ciò consente anche di sviluppare applicativi robot o semi automatici che navigando il web distinguono tali licenze, nelle diverse combinazioni e, operando come se le comprendessero, raccolgono e utilizzano le informazioni disponibili nel rispetto della porzione dei diritti riservati.

Nella recente versione di *Legal Code* espressamente vengono inclusi i diritti *sui generis* connessi ai database. E su questo punto un ulteriore passo in avanti è stato fatto, proprio a parziale copertura dell'incertezza, accennata poc'anzi, nel caso in cui un database dovesse contenere, come di fatto accade, dati pubblici e dati protetti dal diritto d'autore. La protezione piena dei dati coperti dal diritto d'autore non potrà che essere garantita dal costituente della banca dati, mediante appositi sistemi di accesso selettivo sui contenuti, secondo il tipo di licenza acquisita dall'utente.

In definitiva, se sul problema del diritto morale d'autore c'è una visione di salvaguardia pressoché unanime, il nodo che rimane discusso è relativo alle questioni economiche; laddove non si possano diffondere liberamente i dati, una strada che mi sento di condividere, invero già avviata, è quella di incentivare politiche volte ad abbassare le barriere di accesso ai contenuti protetti, ossia renderli più accessibili anche in termini economici e, allo stesso tempo, promuovere l'ingresso ai collettori di informazioni, distinti per tipologie, affinché aggiungano alla base informativa ulteriori servizi in favore dell'utente, in modo da incentivarne l'uso, e prevedano un ragionevole contributo forfettario (o a consumo) in denaro che potrà essere ridistribuito agli autori sulla base delle statistiche di visualizzazione

31 Alle opere così licenziate viene assegnato un codice in HTML che rende riconoscibile a tutti il tipo di protezione optata dall'autore, secondo la combinazione prescelta: BY (Attribution), BY-SA (Attribution-ShareAlike), BY-NC-SA (Attribution-NonCommercial-ShareAlike), BY-NC (Attribution-NonCommercial), BY-ND (Attribution-NoDerivs), BY-NC-ND (Attribution-NonCommercial-NoDerivs). L'inserimento di tali informazioni elettroniche trovano legittimazione nel disposto di cui all'art. 102-quinquies della L.d.A.

32 Per maggiori dettagli sui termini delle licenze, si veda Peruginelli 2007, 9 e ss. Il *Legal Code* 4.0 è pubblicato al seguente indirizzo URL <https://creativecommons.org/licenses/by/4.0/legalcode>.

dei singoli contributi. In altri termini, una parte considerevole degli utenti, che in genere non paga volentieri l'acquisto di singoli dati, se trovasse contenitori coerenti, ragionati e aggiornati di dati di proprio interesse, probabilmente sarebbe più disposta a versare una quota in denaro per poterne fruire in maniera piena e senza limiti, utilizzando i dati nel rispetto del *fair use*, piuttosto che violare le norme e i dispositivi di protezione di volta in volta approntati. Una comunità solidale si sostiene, si sviluppa, progredisce ed è questa la via su cui, più di ogni altra, si scommetterà in futuro.

Per completare il quadro degli aspetti economici legati alla produzione di banche dati, nel senso più ampio, un elemento, non proprio di dettaglio, persiste, oggi più che mai. Mi riferisco alla scarsità di risorse economiche che spesso chi fa ricerca e sviluppo lamenta.

Mi pare utile accennare, ma solo in punta di penna, alle prospettive nuove e interessanti che si stanno consolidando, anche presso le comunità scientifiche, con riferimento alle raccolte di fondi dal basso per la realizzazione di progetti comuni o da integrare ad altri già esistenti. Non sono pochi i denari, per dirla in maniera ortodossa, che è possibile raccogliere mediante piattaforme dedicate al *crowdfunding*, di cui possono beneficiare gli stessi ricercatori per finanziare – sia reciprocamente sia con l'aiuto dei privati, degli operatori economici o di altri enti pubblici, che per qualunque motivo hanno interesse allo sviluppo di un prodotto della ricerca – i loro progetti, purché siano ben presentati e abbiano spendibilità e *appeal* tali da scatenare la corsa alla colletta.

Bibliografia

- Agenzia per l'Italia digitale (2014). *Linee guida per la valorizzazione del patrimonio informativo pubblico* [online]. URL <https://goo.gl/44ZSWo> (2017-10-24).
- Andreoli-Versbacha, Patrick; Mueller-Langer, Frank (2014). «Open Access to Data. An Ideal Professed But not Practised». *Research Policy*, 43(9), 1621-33.
- Assante, Ernesto; Fagnocchi, Marco; Longo, Alessandro (2014). «I pirati che hanno salvato il cinema» [online]. *Le Inchieste*. URL <https://goo.gl/9KzXrh>(2017-10-24).
- Dal Poggetto, Patrizia (2007). «La protezione giuridica delle banche dati mediante il diritto d'autore ed il diritto sui generis». *Informatica e Diritto*, 1, 159-68.
- De Robbio, Antonella (2002). *La tutela giuridica delle banche dati nel diritto d'autore e nei diritti connessi* [online]. URL <http://eprints.rclis.org/4012/1/dbthesis.pdf> (2017-10-24).

- Dominici, Gianni (2014). *Per una seconda fase degli Open Data in Italia* [online]. URL <https://goo.gl/yvRQa3> (2017-10-24).
- Donovan, James; Osborne, Caroline; Watson, Carol (2015). «The Open Access Advantage for American Law Reviews». *Journal of the Patent and Trademark Office Society*, 97(1), 4-25.
- Guercio, Maria (2009). «Archivi digitali. Principi, metodi e criticità organizzative» [online]. *XXI Secolo*. URL [http://www.treccani.it/enciclopedia/archivi-digitali_\(XXI_Secolo\)/](http://www.treccani.it/enciclopedia/archivi-digitali_(XXI_Secolo)/) (2017-10-24).
- Lisi, Andrea; Ungaro, Sarah (2014). *Questioni irrisolte e problemi aperti dal DPCM 3 dicembre 2013 sulla conservazione digitale dei documenti* [online]. URL <https://goo.gl/e6lyZH> (2017-10-24).
- Minazzi, Francesco (2014). *Processo Civile Telematico. I problemi ancora irrisolti nel deposito degli atti* [online]. URL <https://goo.gl/eSLiAC> (2017-10-24).
- Peruginelli, Ginevra (2006). «Problematiche in tema di diritto d'autore e biblioteche digitali». Tammaro, Anna Maria (a cura di), *Biblioteche digitali in Italia. Scenari, utenti, staff e sistemi informativi* [online]. URL <https://goo.gl/2nXkQA> (2017-10-24).
- Peruginelli, Ginevra (2007). *Le licenze Creative Commons. Aspetti critici* [online]. URL <http://www.ittig.cnr.it/Ricerca/Testi/peruginelli2007.pdf> (2017-10-24)
- Scorza, Guido (2015). *Processo civile telematico. Il Parlamento rinvuole la carta. Alla faccia della digitalizzazione!* [online]. URL <https://goo.gl/CdEfbx> (2017-10-24).
- Spampinato, Daria; Zangara, Ignazio (2013). «Classical Antiquity and Semantic Content Management on Linked Open Data». *DH-CASE '13 Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment = Metadata, Vocabularies and Techniques in the Digital Humanities* (Florence, 10th September 2013). Article nr. 13. New York: ACM, 1-7.

The Tesserae Project

Detecting Intertextuality of Meaning and Sound

Neil Coffee

(State University of New York, Buffalo, USA)

Christopher Forstall

(State University of New York, Buffalo, USA)

James Gawley

(State University of New York, Buffalo, USA)

Abstract The Tesserae Project offers a free online intertextual search tool for ancient Greek, Latin, and English. Tesserae has in the past allowed for a pairwise searching of literary texts in these languages for exact word or lemma similarities. This paper describes two new types of search now offered by Tesserae, by meaning (semantic search) and by sound.

Summary 1 Semantic matching. – 2 Sound matching. – 3 Similarity of Meaning and Sound in Intertextual Research.

Keywords Intertextuality. Text reuse. Allusion. Greek literature. Latin literature. English literature. Digital humanities.

The Tesserae Project provides a free online tool that allows users to compare two texts in ancient Greek, Latin, or English and automatically discover instances of similar language, and so identify textual phenomena including text reuse and literary allusion (<http://tesserae.caset.buffalo.edu>). The basic Tesserae search finds sentences or poetic verse lines in two different texts that share two or more similar lemmata. The lemma search is designed to capture parallels where inflectional forms differ, so that 'animo... magno' in one text would match 'animus... magnus' in another. This extends to irregular and suppletive systems so that 'gladium... fert' would match 'gladium... tulit'. The lemma-based search, documented in the project's previous publications, has been shown to capture some 60% of significant parallels found by commentators.

The project's current work involves the development of new search features, with the aim of creating a blended search that better replicates and extends human recognition of textual similarity. These forms of search remain experimental, in that they have not yet been tested and tuned. Nevertheless, they provide new capacities that users can currently employ

Antichistica 14

DOI 10.14277/6969-182-9/ANT-14-14 | Submission 2017-08-14

ISBN [ebook] 978-88-6969-182-9 | ISBN [print] 978-88-6969-183-6

© 2017 |  Creative Commons Attribution 4.0 International Public License

and explore. This article will briefly describe and give examples of two new forms of search, so that users can experiment with them in an informed way. Those who wish to explore the code directly can do so at <https://github.com/tesseract>.

1 Semantic matching

One of the new Tesseract capabilities is the ability to find passages sharing words with similar meanings, regardless of the lemma. This 'semantic search' is available for both the Latin and Greek corpora and adds the ability to compare Latin works directly to Greek. Users can conduct a semantic search from the search page by clicking 'show advanced', then choosing 'semantic' from the feature drop-down list.

Semantic search captures instances of lemma identity, but also parallels that cannot be found through lemma matching. For example, if we compare the whole of Vergil's *Aeneid* with the whole of Ovid's *Metamorphoses*, among the top twenty results is the match of *Aeneid* 3.632, "saniem eructans", with *Metamorphoses* 4.494, "saniemque vomunt". Though Vergil writes of 'belching up gore' and Ovid writes 'they throw up gore', the passage is detected as a potentially significant parallel because 'belch' and 'throw up' are near synonyms. Semantic search works similarly in matching Latin to Greek. In a comparison of all of Homer's *Odyssey* with all of Vergil's *Aeneid*, the phrase κύματ' ἔταμνεν at *Odyssey* 13.88 is identified with "alta secans fluctuque" at *Aeneid* 10.687, where both phrases refer to ships cutting through waves.

As in the lemma-based search, semantic search compares the two selected texts word-by-word. Sentences (or verse lines) sharing two or more semantically-related words are returned as a match. Words are considered related if they appear in the Tesseract related-word dictionary for the language pair being searched - Latin to Latin, Greek to Greek and Latin to Greek. These dictionaries were created by automatically associating the English definitions in lexica retrieved from the Perseus Digital Library: Lewis and Short for Latin and Liddell-Scott-Jones for Greek. Dictionary headwords were compared in a three-step process. First, the most semantically significant words within the definition for each dictionary headword were identified by using TF-IDF scores. Next, these TF-IDF scores were fed into a latent semantic indexing model in order to give each definition a numeric vector representative of its meaning. Finally, the resulting vector-representations of each definition were compared according to their cosine similarity. Pairs of headwords whose entries exceeded a certain threshold were considered 'semantically similar'. Empirical testing of the optimal threshold is ongoing. This method does not distinguish between forms of relatedness: related words could be synonyms, antonyms, or metonyms,

and some are false positives. Preliminary results for the Latin-Greek dictionary suggest, however, that approximately 80% of the correspondences it yields consist of synonyms or closely related words, as in the examples above.

2 Sound matching

Tesserae also has a sound matching capability, using orthography as a proxy for phonology. It is designed to capture instances where one author echoes not the words or meaning of another, but a pattern of sounds. Users can conduct a sound search from the search page for any language, by clicking 'show advanced', then choosing 'sound' from the feature drop-down list.

In a search for sound similarities between Vergil's *Aeneid* and Ovid's *Amores*, among the top results was the line "ipsis praecipuos ductoribus addit honores" at *Aeneid* 5.249 matching the line "haec est praecipuo victoria digna triumpho" at *Amores* 2.12.5. These lines would not have matched by bigram lemma search, since they share only one common word, "praecipuo(s)". They were returned by the sound search because of the additional similarity of the sounds in "ductoribus" and "victoria". The lines deal with related themes. In the *Aeneid*, Aeneas is giving honours to the competitors in the naval race. In the *Amores*, Ovid is celebrating his conquest of Corinna in terms of a military triumph. Both are scenes of military-style competition, victory and celebration. The absence of more than one common word between the two passages makes it difficult to see Ovid's words as an allusion to the *Aeneid*'s line but he seems to be appropriating Vergil's language, sound, and metrical rhythm as ready-made materials to lend a Vergilian feeling to a version of the lover-as-soldier trope.

There are many ways that a sound search could be constructed. Tesserae's sound detection searches for identical three-letter sequences, or character trigrams. The more trigrams shared by two sentences or lines, the higher that pair ranks. Forms from the same lemma are given no special weight, but appear in the results to the extent that they have similar character trigrams.

3 Similarity of Meaning and Sound in Intertextual Research

Traditionally, most instances of Greek and Latin intertextuality documented by scholars have consisted of either quotation or close verbal echo. Computational methods can detect these forms of intertextuality, but have the potential to identify more subtle phenomena as well. Scholars have also long considered that similarities of meaning and sound can contribute

to making a meaningful intertextual parallel, as when they are combined with the reuse of a single lemma, or even constitute such a parallel in the absence of any shared lemmata. The computational capacity to detect semantic and sound similarity across large corpora holds the potential to expand our understanding of intertextuality, allowing us to give fuller consideration to instances where the parallelism between texts is more subtle, but nevertheless significant. We hope that the meaning and sound matching capabilities of the Tesseract website will contribute to advancing this research.

Bibliography

- Coffee, Neil et al. (2012). "Intertextuality in the Digital Age". *TAPA*, 142(2), 381-419.
- Buck, Thomas et al. (2014). "Modeling the Scholars. Detecting Intertextuality through Enhanced Word-Level N-Gram Matching" [online]. *Literary and Linguistic Computing*. URL <https://academic.oup.com/dsh/article-lookup/doi/10.1093/lc/fqu014> (2017-11-13).
- Forstall, Christopher; Scheirer, Walter (2010). "Features from Frequency. Authorship and Stylistic Analysis Using Repetitive Sound". *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(2), 1-23.

Parte 4.

Una Tavola rotonda:

verso la biblioteca digitale antichistica

Introduzione

Anna Maria Tammaro

(Università degli Studi di Parma, Italia)

In questa Tavola rotonda ci si è proposto di iniziare una conversazione tra bibliotecari digitali e umanisti informatici nel settore d'antichistica, che da molto tempo sono impegnati in studi e ricerche sui testi digitali. Hanno partecipato: dalla parte delle biblioteche, Anna Lucarelli (Biblioteca Nazionale Centrale di Firenze); dalla parte degli umanisti informatici, Raul Mordenti (Università degli Studi di Roma Tor Vergata), Maurizio Lana (Università degli Studi del Piemonte Orientale), Raffaele Perrelli (Università della Calabria), Emiliano Degl'Innocenti (SISMEL).

Lo scopo che ci si proponeva era ambizioso: fare un passo avanti nella collaborazione tra bibliotecari e studiosi, anzi di più - cominciare a pensare insieme a un progetto comune per la costruzione della Biblioteca digitale di antichistica. Siamo stati stimolati ad essere ambiziosi dal luogo prestigioso in cui l'Università Ca' Foscari Venezia ci ha ospitato, e anche dall'opportunità di avere per la prima volta insieme i pionieri di biblioteche digitali e i pionieri di informatica umanistica in Italia. C'era però solo un'ora di discussione: questo è stato un limite, anche perché alla fine del Convegno i partecipanti cominciavano a partire.

Malgrado questo limite, alcuni spunti problematici e possibilità di collaborazione sono stati evidenziati con chiarezza, e si possono sommariamente indicare - come verrà precisato nelle conclusioni - come collaborazione delle biblioteche digitali con gli umanisti informatici e collaborazione degli umanisti informatici con le biblioteche digitali.

Biblioteche digitali e umanisti informatici: nuove partnership?

Anna Maria Tammaro
(Università degli Studi di Parma, Italia)

Abstract Until now, we have had insufficient collaboration between digital humanists and cultural heritage institutions for the digitisation of resources for research. In the digital library context, scholars have now had new opportunities to create digital resources tailored to their research needs. The vision of the digital library service is no longer related to an editorial model of digital objects, but rather to a technically possible semantic access model. The Round Table has focused on collaboration on Linked Open Data (LOD), which benefits both scholars and professionals.

Sommario 1 Perché gli umanisti informatici hanno bisogno di una Biblioteca digitale? – 2 Perché le biblioteche digitali hanno bisogno degli umanisti informatici? – 3 Conclusioni.

Keywords Digital Library. Linked Open Data. Scholarly primitives. Ontology. Semantic Web. Open Access. Re-use.

Finora è stato molto insolito che i bibliotecari parlassero di digitalizzazione ai Convegni degli umanisti informatici, come anche finora non c'è stata molta collaborazione tra bibliotecari, archivisti, curatori delle istituzioni culturali e docenti delle facoltà umanistiche per la digitalizzazione del patrimonio culturale. È opinione diffusa che i professionisti, come tecnici della digitalizzazione, possano decidere con quali tecniche e criteri attuare la digitalizzazione meglio anche degli utenti. Si pensa infatti che la digitalizzazione sia di per sé una 'cosa buona' e che, una volta che la collezione digitale sia accessibile, gli utenti sicuramente la useranno, nel modo in cui è stata creata dai professionisti.

La pianificazione della digitalizzazione spesso parte quindi dalla creazione di contenuti, perseguendo un modello editoriale, per poi diffondere le collezioni digitali agli studiosi, istruendoli su come dovrebbero essere usate. Gli studiosi generalmente non hanno nessun controllo dell'edizione scelta e della codifica o struttura della fonte, che vincola le funzionalità disponibili, che quindi possono non essere quelle che servono alla ricerca specialistica. Non si cerca di capire prima come gli umanisti informatici vorranno usare la collezione digitale, ma si pensa spesso a un utente generico.

Non c'è poi da meravigliarsi se si scopre che le risorse digitali create dalle biblioteche digitali, anche investendo notevoli quantità di denaro

pubblico, non sono poi usate. Nella letteratura professionale, quando ci si interroga su questo problema, vengono evidenziati possibili ostacoli all'uso delle risorse digitali come la non conoscenza della risorsa digitale, o la mancanza di capacità digitali per usarla o aspetti di difficoltà di accesso come la mancanza di connessione. Potrebbe invece essere che le risorse digitali non sono usate perché non sono utili agli studiosi? Oppure non consentono il tipo di uso che servirebbe?

Nel contesto della biblioteca digitale, gli studiosi sono stati 'liberati' da questi vincoli e hanno avuto nuove opportunità di scelta di accesso alle risorse digitali, fino alla possibilità di creare proprie risorse digitali adeguate ai loro bisogni di ricerca, anche attraverso il ri-uso di collezioni digitali create da biblioteche. La crescita dei contenuti generati dagli studiosi è particolarmente importante per estendere l'accesso a risorse digitali e a metadati scelti in modo appropriato, anche per dare un valore aggiunto a tali risorse attraverso attività come il collegamento, annotazioni, etc. Possiamo quindi evidenziare numerose iniziative di singoli studiosi o di istituzioni di ricerca che hanno costruito collezioni digitali con le funzionalità necessarie per la ricerca, ma queste iniziative - sempre di grande qualità - sono accomunate da una caratteristica che rischia di togliere valore alle risorse stesse: hanno problemi oggettivi di sostenibilità e permanenza in rete per mancanza di risorse e di competenze tecniche.

Può costruirsi un nuovo rapporto di partnership con le biblioteche digitali su queste esigenze e opportunità?

1 Perché gli umanisti informatici hanno bisogno di una Biblioteca digitale?

La biblioteca digitale di antichistica non si limita a dei contenuti digitali a cui si può accedere in linea per una ricerca veloce. Questa è una concezione della biblioteca digitale che deve essere superata: così è da interpretare il significato del titolo dell'intervento di Maurizio Lana «Oltre le biblioteche digitali». La biblioteca digitale non si limita a memorizzare dei contenuti digitali ma rende disponibili funzionalità avanzate. Il valore della biblioteca digitale è tutto in queste funzionalità avanzate, che giustificano l'investimento notevole che richiede la digitalizzazione, facilitando lo studioso a fare quello che altrimenti non sarebbe stato possibile. Come affermava nel 1962 Douglas Engelbart nell'introduzione di *Augmenting Human Intellect* (citato da Maurizio Lana): «Con 'potenziare l'intelletto umano' intendiamo accrescere la capacità di un uomo di affrontare una situazione che presenta un problema complesso, di acquisire comprensione adatta alle sue specifiche necessità, e derivare soluzioni per i problemi». La 'cura digitale' è la funzionalità di base della biblioteca digitale. A differenza delle risorse digitali accessibili dal web, la biblioteca digitale svolge un

insieme di funzioni, tra cui la preservazione delle risorse digitali raccolte, e questa funzione deve essere apprezzata come la più importante, perché al momento solo la biblioteca digitale assicura la permanenza nel tempo dell'accesso alle risorse. Le altre funzioni che sono comprese nella cura digitale sono l'organizzazione dei contenuti digitali, basata sulla selezione dei contenuti, la struttura del testo e l'immissione dei metadati. La cura digitale richiede un assetto tecnologico, di cui la biblioteca digitale deve dotarsi, che mantiene l'accessibilità alle risorse, malgrado l'obsolescenza del software e la necessità di migrazione dei formati. Le biblioteche digitali sono quindi un assetto digitale molto importante per l'infrastruttura di ricerca (io direi unico) e rappresentano uno spazio virtuale connesso a Internet che mantiene la risorsa accessibile nel tempo per l'accesso, l'uso ed il ri-uso. Ciascuno di questi tre concetti: accesso, uso e ri-uso, devono essere ben compresi e hanno delle problematicità che credo tutti gli studiosi hanno già sperimentato.

Cominciamo con l'accesso. I servizi della biblioteca digitale, oltre a identificare e recuperare la risorsa digitale come in tutti i sistemi di Information Retrieval, offrono una serie di funzionalità avanzate dell'interfaccia. Unsworth (2000) ha chiamato queste funzionalità le «scholarly primitives» o priorità di ricerca che devono essere rese disponibili se si vuole essere di supporto allo studioso: 'Annotating', 'Comparing', 'Referring', 'Selecting', 'Linking', 'Discovery'.

Raul Mordenti nella sua presentazione «Alcuni problemi teorici e pratici posti dalle *digital libraries*» ha evidenziato le difficoltà che ancora si frappongono a un servizio di biblioteca digitale per l'antichistica. Il metodo di ricerca predominante degli umanisti informatici è complesso. Lo studioso ha bisogno di numerose fonti e opere da consultare insieme, e non di eseguire ricerche per parole chiave, mentre la maggior parte invece dei sistemi informativi e delle risorse digitali prevedono questo tipo di ricerca. Lo studioso di antichistica deve seguire le note nel testo che sta leggendo o cercare informazioni collegate in altri testi connessi al primo. Gli umanisti informatici hanno cioè bisogno di fare ciò che Bates chiama 'raccolta di bacche' - in altre parole, lo studioso vuole estrarre citazioni da testi che sono particolarmente attinenti all'argomento di studio, piuttosto che citare tutto quello che è stato scritto sull'argomento (Bates 1989). Hanno anche bisogno di una più ampia gamma di testi, in termini di data di pubblicazione e tipologia di pubblicazioni: invece di leggere articoli di periodici degli ultimi cinque anni, gli umanisti informatici hanno bisogno di consultare libri stampati o manoscritti che hanno centinaia di anni, così come immagini, film, musica, mappe, oggetti museali e altre tipologie di materiale di origine storica. Spesso devono rileggere o riesaminare fonti in un modo complesso, piuttosto che fare la ricerca di documenti digitali per informazioni fattuali.

Un modello di biblioteca digitale basata sulla collaborazione è *digilibLT*, Biblioteca digitale sul latino tardo, presentata da Maurizio Lana nella

Tavola rotonda. La biblioteca digitale *digilibLT* è nata proprio per il ri-uso e distribuisce liberamente sotto licenza Creative Commons le opere del latino tardo in vari formati tra cui ePUB, il formato aperto degli ebook. Il gruppo di lavoro è rappresentato da latinisti insieme a bibliotecari ed informatici.

In conclusione, cercando di rispondere alla domanda: perché gli umanisti informatici hanno bisogno della biblioteca digitale? Sulla base dei progetti sinteticamente descritti dai partecipanti alla Tavola rotonda la prima risposta importante credo che sia 'per la sostenibilità e l'accesso permanente delle risorse digitali'. Un altro motivo è che la biblioteca digitale crea uno spazio virtuale per l'uso, il ri-uso e stimola la collaborazione interdisciplinare. Questa collaborazione ancora non è diffusa, ma viene stimolata dall'Unione Europea e presto sarà comunque evidente dai risultati dei progetti europei e nazionali che sono stati descritti.

C'è un ostacolo però: manca l'accesso aperto alle risorse digitali. I partecipanti alla Tavola rotonda sono stati tutti d'accordo nel definire l'accesso aperto - definito nel modo più ampio - come 'pre-requisito' per la biblioteca digitale d'antichistica.

2 Perché le biblioteche digitali hanno bisogno degli umanisti informatici?

Abbiamo evidenziato come il contesto di riferimento delle biblioteche digitali è diverso da quello delle biblioteche tradizionali e la differenza sta nel ruolo attivo che hanno gli studiosi. Gli umanisti informatici sono creatori essi stessi di contenuti digitali e hanno la possibilità di collaborare alla costruzione e gestione della biblioteca digitale. La prima domanda che era stata fatta ai partecipanti alla Tavola rotonda era: quali sono i contenuti digitali di cui voi siete i creatori e quanto siano *open*, cioè aperti, alla condivisione e collaborazione con altri studiosi anche di discipline diverse?

Un campo di applicazione dove l'apertura sembra avere meno ostacoli è quello dei metadati e delle ontologie. Usando la definizione di Maurizio Lana (citata da Maurizio Lana nel suo intervento durante la Tavola rotonda)

un'ontologia è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di conoscenza, organizzata in una gerarchia di categorie; tale strutturazione permette di costruire ed effettuare per mezzo del computer semplici ragionamenti di tipo formale/deduttivo.

Il Progetto Geolat (Geography for Latin Literature) è un esempio di ontologia che permette a studiosi e cittadini alla ricerca dei nomi geografici contenuti nei testi letterari di scoprire nuovi modi per valutare e leggere i rapporti che legano i testi geografici.

Nel caso dei dati aperti e delle ontologie, la collaborazione tra umanisti informatici e biblioteche digitali diviene necessaria (non solo desiderabile). L'uso dei linguaggi aperti del web è quindi la conseguenza necessaria del bisogno di collaborare.

Annotare i testi utilizzando l'ontologia è un'operazione complessa, che può essere fatta solo da studiosi – avverte Maurizio Lana – perché occorrono elevate conoscenze della lingua e della storia. Ad esempio in Geolat occorre:

- individuare i nomi di luogo (cioè riconoscerli come tali);
- disambiguare le omonimie (per esempio distinguere una dall'altra le varie "Alessandria");
- associare al nome del luogo le caratteristiche appropriate.

Anche Anna Lucarelli è d'accordo sul potenziamento della collaborazione tra bibliotecari e studiosi, partendo dal punto fondamentale che le biblioteche digitali si stanno trasformando da *repositories* di testi a *semantic digital libraries*. Partendo dall'esperienza della trasformazione del Soggettario della Biblioteca Nazionale Centrale di Firenze in relazione alle prospettive introdotte dal web semantico e dai *linked data*, è aumentata la sua portata multidisciplinare. La Biblioteca Nazionale ha la responsabilità istituzionale di mantenere uno strumento di indicizzazione semantico, e il passaggio più opportuno sembra ora quello di avviare una partnership tra il Nuovo Soggettario e gli strumenti di indicizzazione semantica specialistici che sono stati realizzati dagli umanisti informatici, nell'ottica del web semantico.

In conclusione, il progetto cooperativo per una Biblioteca digitale di antichistica oggi non può che passare per i Linked Open Data (LOD) che permettono di collaborare a livello internazionale e interdisciplinare, anche a distanza nel tempo e nello spazio, lasciando a ciascuno dei partner libertà di impianto dei contenuti ma adottando una modalità di interoperabilità molto forte. Per ottenere questo risultato non è sufficiente basarsi sulla tecnologia del web semantico, ma occorre costruire le ontologie su uno schema logico, come ad esempio quello presentato da Marco Giunti per la Biblioteca Iuris Antiqui in questo Convegno. Vi cito questo caso, perché credo che il problema degli Open Data dovrebbe essere al centro della discussione per la costruzione di una Biblioteca digitale di antichistica. Questa discussione dovrà necessariamente avere un focus su uno schema logico, anche se si potranno avere esperienze diverse.

3 Conclusioni

Possiamo concludere dopo la Tavola rotonda del Convegno di Venezia che le funzionalità disponibili nelle attuali biblioteche digitali non sono quelle descritte come prioritarie nell'elenco di Unsworth.

Come potrebbe migliorare la ricerca negli studi di antichistica? La visione del servizio di accesso della biblioteca digitale non è più legata a un modello editoriale di rappresentazione digitale di oggetti in rete, ma piuttosto a un modello di accesso semantico tecnicamente possibile. La base per una biblioteca digitale per l'antichistica sarà costruita dalla forte motivazione a collaborare e a condividere le risorse digitali costruite finora da istituzioni culturali e umanisti informatici, contribuendo in questo modo a una migliore produttività della ricerca.

Sui bisogni di un progetto cooperativo abbiamo ascoltato le relazioni del Convegno e della Tavola rotonda, per concludere che il primo progetto comune di collaborazione dovrebbe essere relativo al ruolo di Linked Open Data (LOD) nello studio del mondo antico. Il termine, LOD comprende approcci alla pubblicazione di risorse digitali che enfatizzano la stabilità, l'accesso granulare al contenuto intellettuale tramite pubblico URI, e la riutilizzabilità come definito mediante licenze che permettono ulteriori copia di materiali disponibili. In una precedente Tavola rotonda sui Linked Data durante l'ultimo Convegno AIUCD, avevo definito i due approcci che si confrontano oggi nel contesto della Biblioteca digitale 'Paradigma chiuso' e 'Paradigma aperto'. I due paradigmi definiscono non tanto la struttura dei dati ma l'apertura alla condivisione e al ri-uso di questi al fine di facilitare la ricerca collaborativa e quindi migliori risultati.

Le biblioteche digitali possono avere un ruolo molto importante nel facilitare la ricerca degli umanisti informatici. Il loro ruolo sta nel promuovere l'accessibilità e la conservazione dei contenuti digitali, attraverso la diffusione dei contenuti creati dagli umanisti informatici, fino alla collaborazione per Linked Open Data e altri strumenti di funzionalità avanzate, come annotare, comparare, connettere, ecc. Mi chiedo perché questa collaborazione ancora non ci sia per le discipline antichistiche, ma sicuramente le tecnologie ci sono, e le metodologie di ricerca e gli strumenti potrebbero essere in grado di adattarsi ai cambiamenti nello scenario in continua evoluzione delle università e della ricerca universitaria.

Ancora oggi, dopo tre anni dalla fine del Convegno, la comunità dei curatori delle istituzioni culturali esperta di Linked Open Data non collabora con la comunità degli umanisti informatici, come ci si proponeva nella conclusione della Tavola rotonda, incoraggiando l'interconnessione e l'interoperabilità delle risorse digitali aperte.

Finora l'adozione di Linked Open Data da parte della comunità degli umanisti informatici è stata infatti modesta. Ancora possiamo parlare solo di 'promesse' di come questa tecnologia potrebbe essere un mezzo per la

pubblicazione e lo scambio di dati tra i progetti degli umanisti informatici e quelli delle istituzioni culturali. Siamo tuttavia ancora convinti che la creazione di ontologie potrebbe diventare un mezzo per condividere concetti astratti complessi e collaborare per approfondire lo studio dei temi di ricerca degli umanisti informatici.

Alcuni degli ostacoli che possiamo elencare circa questa possibile collaborazione tra biblioteche digitali e gli umanisti informatici sono l'insufficienza di strumenti collaborativi a disposizione delle biblioteche digitali, la mancanza di standardizzazione all'interno dei domini di studio degli umanisti informatici, la preoccupazione per l'applicabilità di ontologie specialistiche alle esigenze di un'utenza generica e soprattutto la disponibilità di fonti di dati aperte e 'pulite' cioè rispondenti alle esigenze degli umanisti informatici e con cui iniziare la collaborazione.

Il Convegno di cui si pubblicano gli atti ha avuto il merito di essere il primo sul tema e ha sicuramente aumentato la consapevolezza dei nuovi sviluppi possibili (sia per le risorse digitali che per gli strumenti collaborativi) con la condivisione delle migliori pratiche.

Bibliografia

- Bates, Marcia J. (1989). «The Design of Browsing and Berrypicking Techniques for the Online Search Interface» [online]. URL <https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html> (2017-10-25).
- CENDARI, Collaborative European Digital Archive Infrastructure. URL <http://www.cendari.eu> (2017-10-25).
- COST Action IS1005. «Medieval Europe - Medieval Cultures and Technological Resources» [online]. URL http://www.cost.eu/COST_Actions/isch/Actions/IS1005 (2017-10-25).
- DARIAH, Digital Research Infrastructure for the Arts and Humanities. <http://it.dariah.eu/sito/> (2017-10-25).
- Engelbart, Douglas C. (1962). «Augmenting Human Intellect. A Conceptual Framework» [online]. URL <http://www.dougelbart.org/pubs/augment-3906.html> (2017-10-25).
- TRAME, Text and Manuscript Transmission of the Middle Age in Europe. URL <http://trame.fefonlus.it/trame/index.html> (2017-10-25).
- Unsworth, John (2000). «Scholarly Primitives. What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This» [online]. URL <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html> (2017-10-25).

Oltre le biblioteche digitali

Maurizio Lana

(Università degli Studi del Piemonte Orientale, Italia)

Abstract Geolat (Geography for Latin Literature) is a research project in its startup phase aimed at building a collaborative scholarly reading of classical Latin texts enhancing the reuse of their informative content. The core concept is that the texts are annotated and that the annotation in turn can be used to study the texts. The TEI conforming annotation is aimed to place-names and employs an ad hoc geographical ontology for the world-view of Classical texts. The re-use of the annotation can start from a map interface mixed with a faceted textual search allowing to ask new questions, e.g. which is the geography of Horace's *Odes* and what does it mean? Has it any relation with the geography of other Augustan age authors? Has it any functional or ideological meaning? This basic concept can be expanded adding more ontologies for person names/roles and events/times; or also extending it to more literatures of the European area for an in-depth study of places, roles and events across times.

Sommario 1 La digitalizzazione del patrimonio culturale e la creazione di biblioteche digitali. – 1.1 Segno dell'avanzata pervasiva del digitale. – 1.2 Costosa nell'avvio. – 1.3 Costosa nella manutenzione. – 2 A che cosa serve una biblioteca digitale? – 3 Attività consuete, attività nuove. – 4 Licklider e Engelbart: *Augmenting Human Intellect*. – 5 Dalla ricerca di parole alla ricerca di categorie: l'annotazione semantica. – 6 Un caso concreto: Geolat e la biblioteca digitale digilibLT. – 6.1 digilibLT. – 6.2 Geolat. – 7 Per concludere.

Keywords Digital libraries. Geography. Semantic annotation. Ontologies. Named entities recognition.

1 La digitalizzazione del patrimonio culturale e la creazione di biblioteche digitali

La spinta alla digitalizzazione del patrimonio culturale è continua e pressante. Il patrimonio culturale è delicato e la sua natura fisica ne limita la fruizione a chi si trovi in sua presenza: invece la digitalizzazione prospetta la possibilità che, da un lato, il pubblico fruisca di copie degli originali che rimangono intatti e non soggetti ad usura, dall'altro che molteplici fruitori dispersi nel mondo possano accedere al medesimo bene nella sua forma digitalizzata. Le biblioteche digitali sono un caso paradigmatico di questa tendenza.

1.1 Segno dell'avanzata pervasiva del digitale

Il contesto in cui si colloca la creazione di biblioteche digitali è quello del mondo digitale. Negli anni '90 del secolo scorso l'accesso remoto ai cataloghi delle biblioteche tramite strumenti software di difficile uso (Telnet) appariva un'opportunità incredibilmente avanzata e capace di incidere significativamente nelle pratiche della ricerca in quanto eliminava la necessità di spostarsi fisicamente da una città all'altra per consultare i cataloghi cartacei. Oggi, per chiunque davanti all'esito di una ricerca all'interno di un catalogo bibliotecario c'è l'istinto di cliccare sul titolo per vedere aprirsi in lettura il testo dell'opera - poi la consapevolezza razionale interviene a dire che non sempre questo è possibile. Ma il web ha insegnato questo: che le risorse, i contenuti, i testi (le musiche, i film, le foto) sono a portata di un clic del mouse. La digitalizzazione delle biblioteche rientra in questo grande quadro complessivo di presenza del web nelle vite delle persone e degli studiosi, o per meglio dire di intersezione sempre più inestricabile di mondo fisico e mondo digitale. Poiché parti sempre più ampie e importanti delle vite delle persone si svolgono nel mondo digitale, si tende a portare sempre più nel mondo digitale strumenti e risorse nate nel (per il) mondo fisico.

1.2 Costosa nell'avvio

Restringendo l'attenzione dal tema ampio della digitalizzazione del patrimonio culturale a quello più specifico delle biblioteche e delle biblioteche digitali, occorre in primo luogo notare che i costi da sostenere per digitalizzare il contenuto di una biblioteca sono molto alti, pertanto quasi mai i fondi presenti nel bilancio ordinario sono sufficienti allo scopo e servono quindi finanziamenti ad hoc. Per avere una percezione dell'entità delle cifre in gioco basti pensare al fatto che la digitalizzazione di una parte del posseduto della Biblioteca Nazionale Centrale di Firenze è stata intrapresa da Google per arricchire la biblioteca digitale Google Books - e si sta parlando di una delle più grandi aziende operanti nel mondo digitale (mentre lo Stato italiano non è in condizione di progettare e realizzare in proprio un'iniziativa di questo genere per la conservazione e la diffusione del proprio patrimonio librario). Iniziative più ridotte, mirate ad ambiti disciplinari specifici, sono comunque possibili anche da parte di soggetti più piccoli: per restare in ambito italiano le ben note biblioteche digitali di contenuto latinistico/classicistico Muisque Deoque, digilibLT, BIA, ALIM, CoRiMu, e altre non hanno avuto bisogno per nascere della presenza di grandi aziende del mondo digitale; ma tutte si sono basate (o si basano) su finanziamenti di ricerca specifici.

1.3 Costosa nella manutenzione

Le biblioteche digitali non solo presentano alti costi di avvio ma sono anche molto più delicate e complesse delle biblioteche fisiche. Questa delicatezza si può esemplificare in modo evidente pensando alla sopravvivenza nel tempo: possediamo oggi biblioteche create secoli e secoli fa e siamo in grado di leggere agevolmente le opere che esse contengono; mentre non riusciamo più a leggere il file di un articolo scritto 30 anni fa con WordStar. Si pensi dunque a quanto più delicato sia il funzionamento di una biblioteca digitale che per operare, cioè per mettere le opere a disposizione dei lettori, ha bisogno di un server, cioè di un computer dotato di sistema operativo e programmi, connesso in rete; e ognuno degli elementi dell'insieme (computer, sistema operativo, programmi, connessione in rete) essendo esposto a rapida obsolescenza richiede un costante lavoro di manutenzione e aggiornamento. Ma le parti sono interconnesse e quindi può accadere che l'aggiornamento di una di esse implichi la necessità di aggiornarne altre (e.g. un programma che non funziona più con una nuova versione del sistema operativo) che potrebbero però non essere (più) disponibili in versione compatibile. Poi ci sono i formati dei dati, che evolvono nel tempo e sono anch'essi soggetti a obsolescenza; e i formati di scrittura dei supporti; e gli standard fisici, elettrici, digitali, delle periferiche che leggono i supporti (i connettori degli hard disk ne sono un esempio).

Non essendo praticabile né efficace una politica di conservazione dell'hardware, che ha senso solo in prospettiva museale, si può solo ricorrere a una costosa e impegnativa politica di costante e progressivo aggiornamento evolutivo dell'hardware, del software e dei formati dei dati. Sapendo che ad un certo punto i principi tecnologici su cui si basa la funzionalità della biblioteca si saranno comunque evoluti al punto che essa per continuare a esistere dovrà essere ricostruita da zero - come è ben esemplificato dal 'passaggio' dal motore a vapore al motore a scoppio; o dal motore a scoppio al motore elettrico.

2 A che cosa serve una biblioteca digitale?

Tendiamo a pensare che iniziative di ricerca costose e delicate debbano avere uno scopo forte, che ne giustifichi l'esistenza. La cosa non è ovvia: nelle scienze dure ci sono ambiti di ricerca (la fisica delle particelle, per esempio) che comportano spese immense a tempo indeterminato, in assenza di chiari scopi che configurino una prevedibile utilità di qualche genere per la società che li finanzia. Ciò però non suscita reazioni negative perché in quei campi la ricerca di base con tutti i suoi costi è ammessa senza limitazioni. Le scienze umane si trovano invece a dover giustificare nella società non solo lo scopo e l'utilità delle loro ricerche ma anche la soste-

nibilità economica degli investimenti che richiedono: di qui la domanda sugli scopi di una biblioteca digitale. Di primo acchito si può pensarli come lettura e ricerca di passi. Ma intraprendere la creazione di una biblioteca digitale semplicemente per permettere la lettura delle opere sullo schermo dei computer non sarebbe sostenibile. La disponibilità di opere in formato digitale (digitalizzate) permette però anche la ricerca e individuazione di specifici passi del testo: una semplice ricerca di parole permette di individuare il passo o i passi rilevanti per il proprio scopo, in modo simile a ciò che in passato si faceva consultando le concordanze di un'opera. Le concordanze erano uno strumento di difficile produzione e quindi realizzato solitamente per opere reputate di grande importanza; la digitalizzazione permette lo stesso tipo di lettura analitica per qualsiasi opera disponibile in formato digitale. Ma sono possibili anche altre ricerche più raffinate, come quelle che mirano a individuare la presenza di espressioni di uno scrittore nelle opere di altri scrittori successivi, ricerche in cui l'acume dello studioso viene supportato dalla velocissima, precisa ed esauriente ricerca effettuata dal computer; o quelle che cercano compresenze (co-occorrenze) di parole.

3 Attività consuete, attività nuove

Fin qui vengono realizzate nel mondo digitale le medesime attività che si possono realizzare anche nel mondo fisico con i testi a stampa - ma con maggiore precisione e soprattutto con maggiore velocità, il che significa che ricerche concepibili ma non realizzabili praticamente con le opere a stampa (un caso su tutti: ricerche di co-occorrenze) diventano invece possibili. Ma si possono concepire anche attività nuove, attività che sfruttano a fondo la natura digitale dei testi e li portano al centro dello sviluppo del web del futuro, il cosiddetto 'web semantico'. Perché nel mondo digitale che sta crescendo la cultura classica non deve per forza rimanere nella posizione specialistica/marginale che occupa attualmente nel mondo fisico, continuamente messa alla prova per dimostrare che serve a qualcosa, che ha qualcosa di dire che vale la pena di ascoltare, in quanto appare irrimediabilmente legata al passato - ma può e deve avere uno spazio in cui tutti la possano incontrare come parte del patrimonio culturale complessivo a disposizione nel mondo digitale. Perché sarebbe veramente paradossale che non ci fossero (anche) gli studiosi di scienze umane a far nascere e crescere il web semantico.

4 Licklider e Engelbart: *Augmenting Human Intellect*

Nel 1960 Joseph Carl Robnett Licklider scriveva:

Nella primavera-estate del 1957 [...] cercai di tener traccia di ciò che una persona moderatamente tecnica faceva effettivamente durante le ore che considerava dedicate al lavoro. [...]

Circa l'85 per cento del mio tempo di 'pensiero' era speso nel trovare una posizione per pensare, nel prendere una decisione, nell'imparare qualcosa che sapevo essermi necessario. Molto più tempo veniva speso nel trovare o ottenere informazioni che nello studiarle. Molte ore erano spese nel tracciamento di grafici e altre ore nell'insegnare ad un assistente come tracciarli. Quando i grafici erano finite le relazioni erano ovvie a colpo d'occhio, ma il tracciamento era stato necessario per renderle così evidenti.

Nel corso del periodo che esaminai, insomma, il mio tempo di 'pensiero' era stato dedicato principalmente ad attività che erano essenzialmente esecutive o meccaniche: cercare, calcolare, tracciare, trasformare, determinare le conseguenze logiche o dinamiche di un insieme di assunzioni o ipotesi, preparare la strada per una decisione o una visione approfondita. Per di più le mie scelte di che cosa fare o non fare erano determinate in misura imbarazzantemente grande da considerazioni di fattibilità pratica, e non di capacità intellettuale.

Il principale suggerimento fornito dalle scoperte appena descritte è che le operazioni che riempiono la maggior parte del tempo che si presume dedicato al pensiero tecnico sono operazioni che possono essere eseguite dalle macchine in modo più efficace che dagli uomini. Gravi problemi sono posti dal fatto che queste operazioni devono essere eseguite su differenti variabili e in sequenze impreviste e in costante mutamento. (Licklider 1960, 6)

Senza lasciarsi fuorviare dal fatto che il contesto in cui si colloca il lavoro di Licklider sia di tipo tecnico, il senso complessivo del suo discorso è chiaro e pertinente anche al lavoro di studio in ambito umanistico: il tempo dedicato ad attività eminentemente creative sul piano intellettuale è solo una piccola parte del tempo complessivo speso in una ricerca. La maggior parte è spesa in attività esecutive/ripetitive che però «devono essere eseguite su differenti variabili e in sequenze impreviste e in costante mutamento».

Nel 1962 Douglas Engelbart nell'introduzione di *Augmenting Human Intellect. A Conceptual Framework* scriveva:

Con 'potenziare l'intelletto umano' intendiamo accrescere la capacità di un uomo di affrontare una situazione che presenta un problema complesso, di acquisire comprensione adatta alle sue specifiche necessità, e

derivare soluzioni per i problemi. Capacità accresciuta, in questo senso, è usato per indicare una miscela dei seguenti elementi: comprensione più rapida, comprensione migliore, possibilità di acquisire un livello di comprensione utile in una situazione che prima era troppo complessa, soluzioni più rapide, soluzioni migliori, e la possibilità di trovare soluzioni per problemi che prima sembravano irrisolvibili. E con 'situazioni complesse' indichiamo i problemi professionali dei diplomati, degli uomini d'affari, di chi fa ricerca in sociologia, biologia, fisica, degli legali, dei progettisti – sia che la situazione problematica esista per venti minuti o per venti anni. Non parliamo di trucchi intelligenti e isolati che aiutano in una particolare situazione. Ci riferiamo invece ad un modo di vivere in un dominio integrato dove sensazioni, procedimenti per prove ed errori, intangibles, e la qualità specificamente umana della 'percezione della situazione' coesistono in modo utile con concetti potenti, terminologia e notazione efficace, metodi sofisticati e aiuti elettronici di alto livello. (Engelbart, 1962, 1)

In una sorta di continuazione ideale delle riflessioni di Licklider, Engelbart ritiene che i computer servano al lavoro intellettuale se si riesce a farli essere parte di «dominio integrato dove sensazioni, procedimenti per prove ed errori, e la qualità specificamente umana della 'percezione della situazione' coesistono in modo utile con concetti potenti, terminologia e notazione efficace, metodi sofisticati e aiuti elettronici di alto livello».

Quando un computer dotato di un programma apposito permette, per esempio, di produrre la concordanza di qualsivoglia testo in formato digitale siamo dentro il 'paradigma Licklider': attività esecutive e ripetitive che si basano su un input specifico vengono eseguite da un assistente che le espleta in modo efficientissimo lasciando allo studioso più tempo per ragionare sul significato degli esiti delle ricerche. Il 'paradigma Engelbart' suggerisce qualcosa di più, cioè che il computer diventi parte integrante dell'ambiente di lavoro intellettuale in cui le qualità propriamente umane (sensazioni, procedimenti per prove ed errori, *intangibles*, percezione della situazione) si caratterizzano per essere non-totalmente-formali, il che è un pregio perché permette di operare in modo efficace in tutte quelle situazioni in cui le informazioni, i dati, le conoscenze, non sono completi. In questa situazione il computer e gli esseri umani possono interagire grazie a operazioni esterne e automatiche sui simboli: ciò accade quando i simboli con cui l'essere umano rappresenta i concetti su cui opera possono essere ordinati sotto i suoi occhi, spostati, archiviati, richiamati, modificati, secondo regole anche estremamente complesse, in risposta rapidissima a un minimo di informazione fornita dall'essere umano stesso, per mezzo di speciali dispositivi tecnologici cooperativi:

Nei limiti di ciò che noi possiamo oggi immaginare, potrebbe trattarsi di un computer, con cui potremmo comunicare in modo rapido e facile, accoppiato con un display a colori tridimensionale all'interno del quale potrebbe costruire immagini estremamente sofisticate - con il computer capace di eseguire un'ampia gamma di elaborazioni su parti di quelle immagini, o sulle immagini intere, in risposta automatica a istruzioni umane. (Engelbart 1962, 25)

Engelbart nel 1962 pensava ai simboli come sorte di ideogrammi ma la funzione di quelli che egli chiama simboli la possono anche svolgere le parole, che non rappresentano più se stesse, ma che fungono da etichette, indicatori, che rappresentano gruppi di significati o di concetti - in modo tale per cui gli esseri umani e i computer cooperano nella costruzione di nuova conoscenza.

5 Dalla ricerca di parole alla ricerca di categorie: l'annotazione semantica

Chiunque abbia usato un computer dotato di un programma per ricerca di parole sa bene che se vuole studiare il concetto di abitazione e le forme in cui esso si manifesta nei testi la sua sarà una ricerca di molteplici parole, quale che sia la lingua: per indicare la casa di campagna, la casa piccola, la casa imponente, la casa pubblica, la casa privata, e altro ancora, altrettante e più saranno le parole effettivamente usate nel testo in esame. Meglio sarebbe se tutte le parole afferenti all'ambito lessicale desiderato potessero essere ricercate contemporaneamente passando da una 'analisi del testo quantitativa' a una 'qualitativa', dove 'quantitativo' è ciò che si limita alle stringhe, alle esatte sequenze alfanumeriche da trovare nei testi, mentre 'qualitativo' è ciò che mette in gioco la semantica.

Nel momento in cui in un contesto digitale si introduce la semantica, 'il gioco si fa duro': la semantica, il riconoscimento dei significati (ma si dovrebbe più propriamente dire la corretta classificazione dei significati), la disambiguazione, si ottengono pragmaticamente con simulazioni della conoscenza che un essere umano utilizzerebbe per svolgere quei compiti. Ma pur con le simulazioni, pur con un certo tasso di errore, si possono ottenere risultati interessanti.

Qui si completa il ragionamento partito dalla domanda 'adesso che abbiamo le biblioteche digitali, che cosa ne facciamo?', per la quale le attuali risposte non possono continuare a rimanere valide o per la troppo grande sproporzione fra l'impegno anche economico necessario per realizzarle e il risultato che si ottiene ('leggiamo i testi') o perché non sfruttano appieno le opportunità offerte dal digitale ('facciamo ricerche di parole'). Una delle

possibili risposte adeguate è 'annotiamo i testi per poter effettuare ricerche semantiche', cioè 'portiamo i testi nel web semantico': formulazioni che possono lasciare indifferenti, come se si trattasse di non-risposte alla domanda da cui si è partiti o di discorsi semplicemente tecnici che non toccano lo studioso. In realtà 'portare i testi nel web semantico' significa impiegare ontologie informatiche (esistenti, o da creare ad hoc) per (de) scrivere nel testo in modo formalizzato il senso di parole o proposizioni che esso contiene. Una volta effettuata questa annotazione, essa può essere utilizzata per fare ricerche nel testo per mezzo del computer. Un'ontologia infatti è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di conoscenza, organizzata in una gerarchia di categorie; tale strutturazione permette di costruire ed effettuare per mezzo del computer semplici ragionamenti di tipo formale/deduttivo. Ad esempio se si descrivono per mezzo di un'ontologia gli elementi della geografia fisica (monti, fiumi, mari, territori, ecc.), non solo si indicherà che il Po è un fiume; ma il fiume sarà connesso con altre 'acque dolci' come torrenti; ma anche laghi; e pure stagni; ma più alla lontana anche con altre acque come le 'acque salate' dei mari e degli oceani. Ciò permetterà di ottenere tra gli esiti di una ricerca i passi in cui si menziona il Po sia che si cerchino i 'fiumi', sia che si cerchino le 'acque dolci', sia che si cerchino le 'acque' in generale. E dunque l'annotazione semantica dei testi per mezzo di ontologie è una risposta adeguata alla domanda da cui si è partiti: apre prospettive di studio dei testi molto ampie e innovative, non praticabili senza le biblioteche digitali.

6 Un caso concreto: Geolat e la biblioteca digitale digilibLT

La riflessione teorica sin qui concisamente delineata rimanda in concreto all'attività di ricerca in corso presso il Dipartimento di Studi Umanistici dell'Università degli Studi del Piemonte Orientale con i progetti digilibLT e Geolat. Il primo è stato cofinanziato dalla Regione Piemonte con un bando del 2008 per la ricerca nelle scienze umane e sociali, mentre il secondo è stato finanziato dalla Fondazione Compagnia di San Paolo, in entrambi i casi al termine di procedure di *blind peer evaluation*.

6.1 digilibLT

DigilibLT, Digital Library of Late-Latin Texts (<http://www.digiliblt.uniupo.it>), è un progetto di ricerca di cui è responsabile Raffaella Tabacco, condiretto con chi scrive, per la creazione di una biblioteca digitale del latino tardo, limitata agli autori pagani. La biblioteca raccoglie alla data di composizione di questo scritto 131 opere liberamente disponibili

per la lettura e il download sotto licenza Creative Commons BY-NS-SA, in 4 formati: TXT, PDF, XML/TEI e ePub. Mentre per i formati TXT PDF e XML/TEI le finalità e gli usi sono chiari, può essere interessante segnalare che la disponibilità delle opere anche in formato ePub, unitamente alla possibilità di scaricarle tutte in un unico blocco, permette di portare con sé in un ebook reader tutta la letteratura latina tarda, potendo leggere qualsiasi passo di qualsiasi opera – ciò che praticamente nessuno studioso ha mai potuto realizzare fino ad ora. Per una descrizione ed analisi più dettagliata della biblioteca si veda il contributo di Borgna e Musso in questo stesso volume.

6.2 Geolat

Geolat, Geography for Latin Literature (<http://www.geolat.it>), è un progetto di ricerca di cui è responsabile chi scrive, per l'annotazione semantica dei nomi geografici nella biblioteca digitale digilibLT e in altre biblioteche digitali di testi latini; sui testi così annotati saranno poi possibili ricerche di testo che riguardino specificamente la conoscenza geografica contenuta nei testi stessi.

Il punto di partenza sono le opere degli scrittori latini disponibili in una biblioteca digitale. Si è scelto di partire lavorando sui testi di digilibLT sia per ragioni pratiche sia allo scopo di valorizzare la biblioteca che si sta costruendo, ma – e questo è molto importante – si può operare su qualsiasi opera digitalizzata presente in qualsiasi biblioteca purché la sua licenza lo consenta (occorre una licenza che definisca l'open access alle opere) – l'annotazione infatti può anche essere totalmente esterna al testo annotato, si parla in tal caso di *standoff markup*. In fase iniziale in Geolat i testi vengono copiati dalla fonte (digilibLT) e duplicati internamente, per dar luogo a una annotazione mista, in parte interna ai testi (*inline markup*), in parte esterna. Per avere un'idea di come si può presentare la parte di annotazione inline si veda il metaesempio qui sotto:

```
pax ita conuenerat ut Etruscis Latinisque fluuius <geogName
ref='http://www.geolat.it/geoDat/aTiber' xml:id='tiber01'>
Albula</geogName>, quem nunc <geogName ref='http://www.
geolat.it/geoData/Tiber' xml:id='tiber02'>Tiberim</geogname>
uocant, finis esset
```

Per l'annotazione di passi di testo contenenti riferimenti a luoghi ed entità geografiche lo standard TEI offre un vocabolario piuttosto ricco (capitolo 13 delle TEI Guidelines, «Names, Dates, People, and Places»). Nell'esempio viene usato il termine `geogName`, che “identifica un nome o una frase che si riferisce ad un luogo fisico” per il quale sono utilizzati i seguenti

attributi: @xml:id che identifica univocamente l'elemento annotato e il cui scopo è di permettere rinvii da parte di programmi ai passi di testo che menzionano l'entità geografica; @ref che contiene l'URI che identifica l'entità geografica citata nel passo.

Il dettaglio della descrizione analitica del nome di luogo basata sull'ontologia si collocherà nello standoff markup, utilizzando come connettore tra passo di testo e markup l'attributo 'ref' citato sopra.

L'ontologia permette di descrivere in modo formalizzato i luoghi antichi e la percezione che di essi aveva la cultura del tempo. Essa quindi da un lato contiene la strutturazione concettuale delle entità fisiche (acque, rilievi, e così via) e antropiche (confini, territori, capoluoghi/capitali, ecc.) del mondo antico, ma anche del mondo contemporaneo. Quest'ultimo aspetto potrebbe apparire superfluo o addirittura fuorviante, ma in realtà tenerne conto permette di individuare le corrispondenze e le relazioni tra le entità della geografia antiche e quelle della geografia attuale. Si pensi ai nomi dei monti o dei fiumi, o delle città: occorre poter indicare che la città che si chiamava Augusta Praetoria attualmente si chiama Aosta e continua ad avere la medesima posizione nello spazio; ma dall'altro essa affronta anche questioni complesse e sfaccettate perché occorre rendere conto tra l'altro del fatto che

- le città avevano un fondatore;
- le colonie avevano una madrepatria;
- molti luoghi naturali erano sacri a una qualche divinità;
- le divinità avevano una gerarchia
- dei e dee personificati, principi e qualità deificati (il dio Termine o la dea Giustizia), e altre divinità minori.

E queste caratteristiche possono cambiare da fonte a fonte, o possono presentarsi in modo a loro volta sfaccettato (a seconda dei testi e a seconda del tipo di discorso che viene sviluppato il fondatore di Roma è Romolo oppure anche Enea). Inoltre ci sono aspetti problematici quali:

- la misura dello spazio che ci giunge dal mondo classico è variabile da fonte a fonte e non certa in quanto non si possiedono corrispondenze esatte tra le unità di misura antiche e quelle attuali;
- l'identificazione e il posizionamento esatto dei luoghi possono essere ignoti o dibattuti, come nel caso delle isole Cassiteridi;
- la concezione complessiva dello spazio geografico è differente.

Tali differenze possono dar luogo a problemi irrisolvibili quando si tenti di organizzare e sistematizzare queste conoscenze. Nei testi latini poi si incontrano peculiarità espressive quali, ad esempio, l'indicare un territorio con il nome della popolazione che lo abita - i.e. funge da indicazione geo-

grafica un'espressione che non contiene un nome geografico. E per finire l'ontologia - in quanto viene progettata e costruita anche con finalità di studio (e non solo di descrizione) dei testi annotati - prevede l'esplicitazione di informazioni che si potrebbero definire di supporto, ad esempio

- le determinazioni di tempo relative all'uso di un determinato nome (come nel caso di Felsina/Bononia), o il periodo di tempo in cui un determinato territorio esiste con una identità specifica (la Dacia esiste come provincia a partire dal 102 d.C.);
- l'evoluzione nel tempo di un determinato luogo: Augusta Praetoria prima di essere una città era un castrum;
- o l'indicazione della fonte di autorità su cui si basa una determinata caratteristica associata al nome di luogo: ad esempio l'identificazione delle isole Cassiteridi con le isole Scilly viene completata dalla citazione bibliografica della fonte di tale identificazione; e se vi sono molteplici ipotesi più d'una può essere annotata.

Tali informazioni permettono poi in fase di ricerca sui testi di operare impiegando filtri specifici relativi al periodo in cui una data entità esisteva, o si presentava con determinate caratteristiche, o relativi alla fonte di informazione. In sostanza, ancora una volta, anche nel caso dell'ontologia vale il principio per cui lo strumento sarà tanto più utile quanto più si sarà progettato fin dall'inizio l'uso che se ne vuole fare o che se ne potrà fare.

L'ontologia geografica è accompagnata da un'ontologia testuale (Open Annotation), che permette di descrivere le relazioni tra autore, testo, suddivisioni del testo, passi, varianti, necessaria perché ogni entità geografica, man mano che le sue occorrenze vengono annotate nei testi, si arricchisce dell'elenco dei passi stessi. In tal modo l'ontologia non rimane una strutturazione della conoscenza, ma diventa una raccolta ordinata, enciclopedica, di informazioni.

Annotare i testi utilizzando l'ontologia è un'operazione complessa, che può essere fatta solo da studiosi perché occorre

- individuare i nomi di luogo (cioè riconoscerli come tali);
- disambiguare le omonimie (per esempio distinguere una dall'altra le varie "Alessandria");
- associare al nome del luogo le caratteristiche appropriate.

Tutte operazioni che richiedono solide conoscenze di lingua e cultura classiche. Nello stesso tempo è chiaro che servono anche strumenti per sveltire le operazioni di annotazione e a questo scopo si ricorre a un ambiente di lavoro specificamente destinato all'annotazione dei testi e a strumenti di lemmatizzazione integrati da procedure di Named Entities Recognition.

L'ambiente di lavoro permette di sveltire le operazioni ripetitive: se un

nome di luogo presente in un dato passo è già stato annotato in precedenza in un altro passo di un'altra opera, l'ambiente di lavoro lo segnala, lo studioso verifica che non si tratti di un'omonimia, e associa al passo di testo in esame la descrizione del nome di luogo già precedentemente inserita. La lemmatizzazione (che si realizza grazie alla collaborazione dell'Istituto di Linguistica Computazionale del CNR) e la NER permettono di individuare i nomi geografici, di identificarli, e di associare a essi un'annotazione minima che contiene alcune caratteristiche descrittive di base: le coordinate GPS e le traduzioni del nome nelle lingue moderne, ove disponibili. Lemmatizzazione e NER sono soggette a errori, tanto più sui nomi propri, ma forniscono un grande aiuto nell'annotazione dei nomi più frequenti, e permettono che il lavoro dello studioso si concentri sugli aspetti più creativi.

Ovviamente non è pensabile che un piccolo gruppo di persone possa provvedere a una così complessa annotazione dei nomi geografici in tutta la letteratura latina tardoantica in pochi mesi. L'intero sistema e ambiente di lavoro è pertanto concepito in modo da permettere la collaborazione di molteplici studiosi nel corso del tempo – perché in certa misura l'annotazione dei testi può essere concepita come un modo per lo studioso per riportare in contesto l'esito di uno studio. Se Mario Rossi pubblica un articolo in cui propone una nuova identificazione delle isole Cassiteridi, egli può annotare l'entità geografica 'isole Cassiteridi' in Geolat inserendo la sua identificazione nella geografia contemporanea, completa di citazione della fonte. Il testo antico diventa così contenitore della conoscenza che si possiede su di esso, in modo non molto dissimile da quanto avveniva con gli scholia.

Tutto questo complesso lavoro permette poi di costruire un'interfaccia di ricerca e consultazione dei testi geograficamente potenziata. Nella sua forma più estrema questo significa che sarà possibile fare ricerche sui testi partendo da una carta geografica sulla quale si traccia un cerchio, ottenendo in risposta l'elenco e la visualizzazione dei nomi di luogo di tale area presenti nelle opere e negli autori annotati, oltre all'elenco delle opere e degli autori che menzionano luoghi collocati in quell'area (ovviamente dagli elenchi dei passi si potrà andare alla loro lettura). Ma l'ontologia permette di creare ricerche o di filtrare quelle già fatte: 'la categoria delle acque negli scrittori del III e IV secolo d.C.'; oppure di creare ricerche miste che integrano la normale ricerca di parole con la ricerca di categorie dell'ontologia: 'la categoria dei rilievi e la menzione dei serpenti all'interno del medesimo capitolo nelle opere di genere storico'. In altre parole Geolat offre un'interfaccia di ricerca che aggiunge capacità geografiche a quelle consuete (ricerca di parole, generazione di concordanze, e simili); e le capacità geografiche hanno un certo grado di finezza semantica grazie all'impiego dell'ontologia.

L'interfaccia di ricerca qui rapidamente delineata è solo uno dei possibili

impieghi dell'annotazione del testo per mezzo dell'ontologia geografica. Infatti l'annotazione del testo sarà pubblicata sotto forma di dati aperti che qualsiasi servizio digitale potrà interrogare e riusare; per esempio un soggetto che opera nel mondo del turismo potrà mettere a punto la sua offerta cercando quali località antiche menzionate nei testi si trovano in una determinata area e individuandone i corrispondenti nomi attuali nell'ontologia di Geolat e nei testi latini; analogamente potrebbe fare un individuo che voglia costruire un viaggio sulle tracce della 'Peregrinatio Egeriae'; e analogamente potrebbe fare una persona semplicemente curiosa così come già accade con le mappe di Google: si clicca, si fa zooma, ci si sposta – come dire che il mondo antico è presente, è a portata di clic, ci si può curiosare dentro con lo stesso atteggiamento con cui si curiosava nel mondo di oggi con Google Maps.

7 Per concludere

Quello che è stato qui velocemente delineato è un discorso molto complesso, in cui si connettono – al di là di una loro esplicita e consapevole intenzionalità – la ricerca e i suoi sviluppi, le finalità degli studiosi, l'interesse pubblico, gli sviluppi del web, lo spazio che negli anni a venire gli studi classici potranno avere o sapranno conquistarsi nel discorso pubblico che si volgerà nel mondo digitale. Fino a qualche anno fa gli studiosi potevano ignorare tutto ciò, concentrandosi sul fare buona ricerca, ma oggi non è più possibile. Una parte crescente delle fatiche degli studiosi va nella ricerca di finanziamenti, quei finanziamenti per i quali, sempre in misura più o meno grande, viene chiesto quale sia l'interesse pubblico a che quella data ricerca si realizzi – segno che le nostre discipline devono diventare capaci di raccontare in modo convincente nella società i loro scopi e i loro sogni, perché 'il classico' non può più far conto su un prestigio e un'autorevolezza riconosciute a priori.

Bibliografia

- Ciotti, Fabio; Lana, Maurizio; Tomasi, Francesca (2015). «TEI, Ontologies, Linked Open Data. Geolat and Beyond» [online]. *Journal Of The Text Encoding Initiative*. DOI 10.4000/jtei.1365.
- Corcione, Claudia et al. (2015). «GO! Geographical Ontology» [online]. URL <https://goo.gl/2WmQH8> (2017-11-13).
- Engelbart, Douglas (1962). *Augmenting Human Intellect. A Conceptual Framework* [online]. SRI Summary Report AFOSR-3223. URL <https://goo.gl/n5Y11E> (2017-11-13).
- Lana, Maurizio (2015). «Biblioteche digitali e Linked Open Data. Contenuti, connessioni, link, argomentazioni». *Digitalia*, 10, 20-9.
- Licklider, Joseph (1960). «IRE Transactions on Human Factors in Electronics» [online]. URL <https://goo.gl/vndAUs> (2017-11-13).

Biblioteche digitali e specialisti del mondo antico Collaborazioni nell'ambito dell'indicizzazione e della ricerca semantica

Anna Lucarelli

(Biblioteca Nazionale Centrale di Firenze, Italia)

Abstract National libraries can play an important role in encouraging the development of digital libraries in classical studies. Reciprocally, the community of classical studies scholars could enhance their collaboration with these libraries to supplement and enrich existing resources. For this purpose, a good starting point would be the subject indexing tool edited by the National Library of Florence: a multidisciplinary Thesaurus whose metadata are available in formats compatible with the semantic web. For many years the Nuovo soggettario terminology has increased thanks to the proposal of experts and researchers in specialised fields. The collaboration between librarians and researchers and the resulting melting pot of their knowledge allow for the construction of tools useful for the common benefit of the next generation of the digital libraries.

Sommario 1 Il contesto: punti critici e nuove possibilità. – 2 Il Thesaurus del Nuovo soggettario come strumento per la collaborazione in ambito antichistico.

Keywords Classical studies. Digital libraries. National libraries. Nuovo soggettario. Subject indexing. Thesauri. Semantic web.

1 Il contesto: punti critici e nuove possibilità

Il rapporto fra il mondo della ricerca e le biblioteche dovrebbe essere particolarmente attivo e fruttuoso quando si parla di strumenti di recupero dell'informazione e in particolare di indicizzazione semantica, un'attività che sta acquisendo un interesse sempre maggiore in relazione alle biblioteche digitali e allo sviluppo del 'web dei dati'. Sinora, un colloquio attivo fra esperti di particolari discipline e bibliotecari che operano per la messa a disposizione e l'aggiornamento di strumenti di indicizzazione efficaci e condivisibili ha stentato a decollare con decisione, pur non mancando, anche in ambito italiano, esperienze prototipali.

Nel contesto di questa Tavola rotonda, è interessante confrontarsi sul ruolo che l'allestimento di strumenti di indicizzazione orientati al web semantico può svolgere per favorire lo sviluppo delle biblioteche digitali e, più in particolare, su quanto le Biblioteche nazionali possono contribu-

ire a migliorare i risultati della ricerca semantica (come è noto, basata sui contenuti concettuali delle opere descritte, analizzati e sommarizzati tramite linguaggi documentari uniformi e standardizzati) a favore di studiosi e ricercatori di Antichistica. Come sappiamo, in quasi tutti i paesi del mondo sono proprio le Biblioteche nazionali ad avere il compito istituzionale di curare e sviluppare gli strumenti di indicizzazione impiegabili trasversalmente, a vari livelli, sull'intero territorio. Strumenti che, peraltro, pur nascendo nell'alveo di biblioteche, trovano spesso una diffusione più ampia nella soggettazione di risorse non solo bibliografiche, ma anche archivistiche, multimediali, grafiche, ecc.

Nello stesso tempo, dovremmo anche discutere di come, inversamente, una specifica comunità di studiosi, come quella che si è riunita per questo convegno, possa collaborare con biblioteche anche generaliste (più scontata la possibilità di un loro apporto a quelle accademiche e specializzate) per implementare, arricchire, integrare e rendere aperte e fruibili le risorse digitali, nuove o già esistenti, da queste prodotte. Come ha già detto nella sua introduzione Anna Maria Tammaro, molto più di quanto già avvenga, le biblioteche potrebbero raccogliere i frutti del lavoro di questi specialisti, non solo in termini di contenuti documentali e testuali, di edizioni critiche, ma anche di dati e metadati. Un punto fondamentale da cui partire è il fatto che le biblioteche digitali si stanno trasformando da *repositories* di testi a *semantic digital libraries*, i cui contenuti - soprattutto per quelle di ultima generazione - potranno senza dubbio arricchirsi da un potenziamento della collaborazione e del colloquio con utenti/studiosi (cf. Biagetti 2014).

In quest'ottica, ed è forte allora il riferimento ai contributi presentati nella sessione mattutina di questo convegno dedicata agli «Strumenti di ricerca semantica», potremmo chiederci se gli strumenti di indicizzazione già realizzati e disponibili, in Italia ma non solo, possano costituire un buon punto di partenza, un terreno in cui sperimentare la partnership citata: una piattaforma da cui far decollare dialogo, collaborazioni, convergenze di risorse intellettuali e tecnologiche.

Nell'ottica dei rapporti fra strumenti di indicizzazione semantica e creazione di biblioteche digitali di Antichistica, a mio avviso, potrebbero essere assunte tre prospettive, così schematizzabili nel modo necessariamente sintetico che questa sede impone.

La prima: quali attori per quale progetto? Mi piace pensare che una partnership con il mondo della ricerca dovrebbe veder coinvolte in un piano coordinato non solo le biblioteche, ma anche gli archivi e i musei. L'idea che non si possano più pensare come entità contenutisticamente e tecnologicamente separate le diverse istituzioni depositarie dei dati della memoria culturale di un Paese o di una comunità è ormai consolidata a livello internazionale e testimoniata da un numero considerevole di iniziative, progetti, portali, ecc. Questa visione di biblioteca digitale come

infrastruttura di contenuti misti pervade, solo per fare qualche esempio, esperienze ben note come Europeana;¹ come GLAM (acronimo per Galleries, Libraries, Archives, and Museums), nato nell'ambito della galassia wikipediana;² oppure come il tedesco BAM, Internet-Portal für Bibliotheken Archive Museen.³

La seconda: come rafforzare e condividere anche in Italia esperienze straniere di collaborazione fra bibliotecari e specialisti di Antichistica? Penso a quanto realizzato, ad esempio, nell'ambito di progetti prestigiosi come Perseus Digital Library,⁴ descritto durante questo convegno da Gregory Crane, oppure come Biblissima, il cosiddetto 'observatoire du patrimoine écrit du Moyen Age et de la Renaissance', allestito in Francia grazie a un programma cooperativo fra più istituzioni e al quale dà un contributo sostanziale il Dipartimento Manoscritti della Bibliothèque Nationale.⁵

La terza: da quali concreti strumenti di indicizzazione già realizzati sarebbe possibile prendere le mosse per implementare, in modo collaborativo, biblioteche digitali di Antichistica? In Italia esistono già *tools* utilmente sfruttabili in questo senso. Ne è un esempio il Nuovo soggettario che la Biblioteca Nazionale Centrale di Firenze (BNCF) mette a disposizione di biblioteche digitali, anche specializzate, offrendo contemporaneamente la possibilità di sviluppare strumenti catalografici e di accesso all'informazione con l'apporto di competenze e risorse di varia natura e grazie alla cooperazione fra istituzioni anche molto diverse fra loro.

2 Il Thesaurus del Nuovo soggettario come strumento per la collaborazione in ambito antichistico

Il *Nuovo soggettario* è un sistema di indicizzazione per soggetto allestito per l'intero sistema delle biblioteche italiane e ha un'applicabilità sia in ambienti di pre-coordinazione (orientati alla costruzione di stringhe di sog-

1 URL <http://www.europeana.eu/portal> (2017-10-26).

2 URL <https://it.wikipedia.org/wiki/Progetto:GLAM> (2017-10-26).

3 URL <http://www.landesarchiv-bw.de/web/44573> (2017-10-26). L'IFLA (International Federation of Library Associations and Institutions) ha preso in carico già da alcuni anni questa problematica; si veda anche Tammaro 2014. Su un interessante progetto di interoperabilità fra biblioteche e archivi nella Repubblica Ceca, finalizzato alla creazione di infrastrutture per la condivisione e il riuso di contenuti digitali e di metadati, si veda inoltre il recente contributo presentato alla IFLA Conference di Cape Town (15-21 Agosto 2015): Balíková Marie et al. 2015.

4 URL <http://www.perseus.tufts.edu/hopper> (2017-10-26).

5 URL <http://www.biblissima-condorcet.fr> (2017-10-26).

getto), sia di post-cordinazione (basati sull'attribuzione di parole chiave).⁶ Il lavoro ha seguito varie fasi e sta continuamente evolvendo anche, e soprattutto, in relazione alle prospettive introdotte dal web semantico e dai linked open data.⁷ La sua caratteristica di sistema impiegabile in vari ambiti, a vari livelli, e in vari contesti e domini disciplinari (che lo allinea, da questo punto di vista, ai suoi equivalenti sistemi diffusi in altri Paesi del mondo come Library of Congress Subject Headings, RAMEAU, Schlagwortf normdatei, ecc.), è alimentata dalla portata multidisciplinare della sua componente principale, un Thesaurus online completamente aperto, dal 2013 con interfaccia anche in inglese. Questo vocabolario controllato è allestito secondo principi e standard internazionali ed è ormai ricco di oltre 55.000 termini afferenti a ogni ambito disciplinare, collegati da relazioni semantiche e impiegabili sia in fase di indicizzazione che di ricerca.⁸ La fig. 1 mostra la pagina di ricerca nella versione per l'utente.



Figura 1. Interfaccia di ricerca del Thesaurus del Nuovo soggettario

La sua natura di *general thesaurus* (o, come si preferisce dire in ambiente anglosassone, *universal thesaurus*) non confligge affatto con la possi-

6 Per una descrizione del linguaggio documentario nel suo insieme: Biblioteca Nazionale Centrale di Firenze 2007. Per l'home page del sistema: <http://thes.bncf.firenze.sbn.it/index.html> (2017-10-26).

7 URL <http://linkeddata.org/> (2017-10-26).

8 URL <http://thes.bncf.firenze.sbn.it/ricerca.php> (2017-10-26).

bilità di essere impiegato anche in contesti documentari specialistici; è adattabile a ogni dominio dal punto di vista strutturale e può accogliere terminologia anche molto specifica e settorialmente specializzata, senza particolari limitazioni, se non quella della 'garanzia letteraria', principio cardine di ogni schema o linguaggio di classificazione o di indicizzazione per soggetto. Una problematica, quella del rapporto fra multidisciplinarietà e specializzazione, che nei sistemi di organizzazione e gestione della conoscenza viene, peraltro, sempre più affrontata in una prospettiva dinamica, approccio d'altra parte supportato dalla stessa Antichistica, un caso esemplare di meta-dominio, di universo interdisciplinare che coinvolge competenze diverse.

Il Nuovo soggettoario è un sistema in evoluzione sul piano quantitativo (accrescimento continuo del Thesaurus e dei suoi link); qualitativo (funzionalità logiche e tecniche) e cooperativo (allargamento della rete di istituzioni coinvolte). Accoglie terminologia sia predisposta dalla BNCF, sia proposta da biblioteche, istituzioni culturali e centri di ricerca, in fase di concreta indicizzazione per soggetto di specifiche opere.

Il Thesaurus sviluppa interoperabilità con altre risorse, nell'ottica del web semantico. Dal 2010 adotta il formato SKOS (Simple Knowledge Organisation System), un *data model* che assegna un valore univoco ai concetti, indipendentemente dalla loro espressione linguistica/letterale, e che è stato sviluppato dal W3C Semantic Web Deployment Working Group (SWDVG) per favorire l'uso, la condivisione e l'interoperabilità di sistemi di organizzazione della conoscenza (come thesauri, classificazioni, tassonomie, ecc.) nel contesto del web, anche al di fuori di ecosistemi strettamente bibliografici.⁹ Le biblioteche (non solo le Nazionali) di tutto il mondo stanno progressivamente adottando questo standard per i propri vocabolari controllati, permettendone il collegamento con universi digitali che rimarrebbero altrimenti separati. I metadati del Thesaurus del Nuovo soggettoario sono disponibili anche nelle triple del linguaggio RDF (*Resource Description Framework*), sono open source, integralmente scaricabili senza registrazione obbligatoria, liberamente usabili secondo la Licenza Creative Commons (CC BY 2.5 IT) e interrogabili tramite un endpoint SPARQL sulla piattaforma DataHub,¹⁰ dove lo strumento viene così affidato alla ribalta dei citati linked open data (cf. anche Bergamin, Lucarelli 2013).

Grazie a questi formati, il Thesaurus colloquia tramite link (basati su relazioni *skos:closeMatch*) con altri strumenti digitali, con altri thesauri e sistemi di classificazione o con enciclopedie. Prevede circa 8.000 col-

9 URL <http://www.w3.org/2004/02/skos> (2017-10-26).

10 Per informazioni più dettagliate, si veda anche la pagina del sito dedicata ai suoi formati e protocolli standard: <http://thes.bncf.firenze.sbn.it/thes-dati.htm> (2017-10-26).

legamenti reciproci agli equivalenti in inglese della Library of Congress (LCSH);¹¹ 3.000 a quelli del sistema di indicizzazione francese RAMEAU; più di 13.000 link reciproci con Wikipedia e con strumenti ben noti come Agrovoc, Eurovoc, con la banca dati DoGi, lo schema di classificazione delle materie giuridiche prodotto dal Consiglio Nazionale delle Ricerche, ecc.

Ma ciò che può essere interessante rilevare in questa sede, in relazione alla possibilità di una sua valenza per migliorare l'efficacia e la funzionalità di biblioteche digitali legate agli studi sul mondo antico, sono due fattori importanti.

Il primo di questi è la grande quantità di terminologia di ambito antichistico che il Thesaurus contiene già: migliaia di termini (riferiti a linguistica, filologia, letterature classiche, archeologia, storia antica, paleografia, diritto romano, ecc.), navigabili fra loro, controllati nelle forme e nei significati, corredati di relazioni semantiche, di fonti repertoriali generali e specialistiche, di equivalenti in altre lingue, di collegamenti alla risorse descritte nel catalogo online della BNCF, ecc. Un patrimonio lessicale che può favorire gli antichisti costituendo l'ossatura di percorsi e di mappe semantiche utili nell'organizzazione di contenuti concettuali e nel collegamento fra biblioteche digitali, *repositories* di altre istituzioni culturali, banche dati di varia natura.

Le figg. 2 e 3 presentano alcuni esempi di termini del Thesaurus, corredati di collegamenti attivati tramite i campi Fonte, oppure Equiv. LCSH.

11 Sui collegamenti reciproci fra Nuovo soggettario e Library of Congress Subject Headings: Lucarelli, Viti 2015.

The screenshot shows the 'Nuovo soggettario THESAURUS' interface. The search results for 'Letteratura bizantina' are displayed. On the left, there is a sidebar with search filters and system information. The main content area shows the term 'Letteratura bizantina' with its classification (UF, TT, BT, RT) and a list of related terms. A definition and source information are provided at the bottom.

Letteratura bizantina RDF/XML

— Ricerca per

- termino
- da term
- categorie/locati
- DDC
- riserve obsolete
- varianti storiche
- equivalenti LCSH

+ Soggettario(1996)
Lista di aggiornamenti (1996-1998)

Sistema Nuovo soggettario

- home page
- manuale applicativo

Letteratura bizantina Cerca Storia Modifiche

Categoria/Fascetta: [Azioni/Atività](#)

UF [Letteratura greca bizantina](#), [Letteratura greca medievale](#)

TT [Atività](#)

BT [Letterature elleniche](#)

RT [Filologia bizantina](#)
[Letteratura greca](#)
[Lingua greca bizantina](#)
[Forme bizantine](#)
[Romanzi bizantini](#)
[Storia bizantina](#)

Definizione: Letteratura greca che vide il regno di Giustiniano (527 d.C.) alla caduta di Costantinopoli (1453 d.C.) (PT)

Fonte: [Soggettario](#) (PT); [SUL](#); [RAVEAU](#); [Librarians](#) (bizantina); [WebDewey](#) (IT); [Wikipedia](#) (IT)

DDC (WebDewey): 883

Equiv. LCSH: [Byzantine literature](#)

Figura 2. Il record del termine Letteratura bizantina nel Thesaurus

The screenshot shows the 'Nuovo soggettario THESAURUS' interface. The search results for 'Coemptio' are displayed. On the left, there is a sidebar with search filters and system information. The main content area shows the term 'Coemptio' with its classification (TT, BT, RT) and a detailed definition. A historical note and source information are provided at the bottom.

Coemptio RDF/XML

+ Ricerca per termine dal Nuovo soggettario

+ Soggettario(1996)
Lista di aggiornamenti (1996-1998)

Sistema Nuovo soggettario

- home page
- manuale applicativo

Coemptio Cerca Storia Modifiche

Categoria/Fascetta: [Cose/Strumenti](#)

Nota d'ambito: Nel diritto romano, in particolare nel periodo arcaico di Roma, uno dei modi mediante i quali la donna entrava a far parte di un'altra famiglia tramite la vendita da parte del pater familias al futuro marito, con la quale la donna veniva assoggettata alla potestà di questo (o del padre di lui, se il marito vi era a sua volta soggetto)

TT [Strumenti](#)

BT [\[Strumenti del diritto romano\]](#)

RT [Confarreatio](#)
[Diritto romano privato](#)
[Matrimonio](#)

Nota storica: Mai usato in BNI fino al 2012

Fonte: [Soggettario](#); [Treccani.it](#) (voce: Matrimonio); [DeM](#); [BIAnet](#); [EDInG](#); [EDRL](#); [EG](#); [Wikipedia](#) (IT)

Associaz. catalografica/Proprietarie: [BN](#); [TTIG](#)

Figura 3. Il record del termine *coemptio* nel Thesaurus

Il secondo fattore da considerare è la natura collaborativa di questo strumento. Come già accennato, esso si sviluppa già grazie a un modello cooperativo basato su una rete di interazioni e scambi. Nella sua home page si legge che partecipano al progetto reti e consorzi di biblioteche come SDIAF (il Sistema documentario integrato dell'Area fiorentina), CoBis (il Coordinamento biblioteche speciali e specialistiche di Torino), la rete delle biblioteche afferenti alla Conferenza episcopale italiana; ma anche biblioteche universitarie (come quelle dell'Università di Bologna, Castellanza, Firenze, Pisa, della Statale di Milano, dell'Università Bocconi), centri interuniversitari, di documentazione e di ricerca (come il CIR di Catania dedicato all'Informatica romanistica, l'ITTIG del CNR), biblioteche specializzate e centri di documentazione di altri enti pubblici o privati (la Biblioteca Centrale Giuridica, il Centro di documentazione di Polimoda), prestigiose istituzioni culturali italiane (l'Accademia della Crusca, l'Istituto della Enciclopedia Italiana), e inoltre, nella citata ottica GLAM, la Soprintendenza Archivistica per la Toscana e alcuni musei. E infine, la fondazione Wikimedia Italia, grazie alla cui partnership viene sviluppata sia l'interoperabilità del Nuovo soggetto con Wikipedia che la collaborazione della BNCf al progetto Wikisource.¹²

Gli enti citati, molti dei quali, peraltro, sono specializzati anche in discipline antichistiche, propongono tramite procedure stabilite e formalizzate nuovi concetti/termini per il Thesaurus e forniscono consulenze scientifiche su significati, morfologie, possibili relazioni semantiche, ecc. Si tratta di una collaborazione che al momento riguarda sostanzialmente bibliotecari (quelli della BNCf e quelli degli enti coinvolti nella cooperazione) e che in qualche caso però ha già visto l'entrata in gioco anche di docenti e studiosi che, coinvolti su specifiche questioni dai bibliotecari delle università di appartenenza, hanno dato consulenze scientifiche di cui il Nuovo soggetto ha potuto beneficiare. Questo scambio di conoscenze e di reciproco arricchimento di competenze si verifica soprattutto quando i contatti della BNCf si stabiliscono con università o istituzioni in cui c'è un rapporto più fervido fra il capitale umano della biblioteca e quello impegnato nella ricerca e nella didattica come, ad esempio, nel caso dell'Università Bocconi, dell'Università di Pisa oppure dell'Accademia della Crusca.

Cosa si potrebbe fare ancora? Un ulteriore coinvolgimento di antichisti nel progetto della BNCf potrebbe portare all'aggiornamento delle fonti attualmente impiegate da chi struttura termini per il Thesaurus,¹³ a proposte

12 Per una sintesi dei comuni progetti fra BNCf e Wikimedia: Lucarelli 2014.

13 Ad oggi, le fonti repertoriali, enciclopediche, ecc. usate sono circa 500, di cui moltissime in formati digitali. Per l'elenco completo: <http://thes.bncf.firenze.sbn.it/fonti.php?menuR=2&menuS=2> (2017-10-26).

di incremento delle voci, al collegamento con risorse digitali e bibliografiche (ad esempio, articoli di riviste) e con archivi di altre parole chiave estratte da testi, alla fornitura di annotazioni o di altre informazioni da segnalare in appositi campi Note. Non ultimo, la collaborazione potrebbe contribuire ad avviare progetti di soggettazione di opere antiche, un'attività che le biblioteche italiane hanno purtroppo da sempre evitato sulla base di pregiudizi infondati, nonostante alcune iniziative ne esprimano periodicamente l'utilità.¹⁴

Un modello in cui la ricerca apporti e restituisca valore aggiunto a quanto allestito e sviluppato dalle biblioteche potrebbe delineare il Nuovo soggettario, basato su standard e tecnologie compatibili con il web semantico, come *tool* di riferimento anche per studiosi di Antichistica, configurandolo come un collettore di altre condivisibili risorse. Sarebbe interessante che altre biblioteche digitali già sviluppate in questo settore, concettualmente connesse ma ancora separate sul piano dell'interoperabilità, potessero sfruttare questo strumento come possibile *hub*, come uno snodo che faciliti la loro integrazione. Significative potenzialità, in questo senso, stanno emergendo, per fare un esempio concreto, dal colloquio che il Thesaurus ha già iniziato a stabilire sperimentalmente con BIA-Net,¹⁵ il progetto dedicato al patrimonio scientifico legato al diritto romano e antico, curato da Nicola Palazzolo, direttore del Centro interuniversitario per l'informatica romanistica di Catania, e le cui evoluzioni sono illustrate proprio nell'ambito di questo convegno sia da Daria Spampinato che da Marco Giunti.

L'interazione fra Nuovo soggettario e BIA-Net sta peraltro configurandosi come un'ottima 'palestra' prototipale soprattutto per la possibilità di integrare fra loro habitat di tipo multidisciplinare e specialistico, oltre che un'interessante esperienza di lavoro comune fra indicizzatori ed esperti di particolari discipline.

Non ci sono dubbi che le biblioteche digitali, in un'ottica rinnovata e incentrata sull'utente oltre che sui contenuti (come vogliono le *semantic digital libraries* di ultima generazione), potrebbero trarre significativi benefici dal potenziamento e da ulteriori sviluppi di iniziative di tal genere, così come da sinergie fra progetti, sistemi di organizzazione della conoscenza, data base collegati a motori di ricerca, strumenti multilingue. Prodotti che esistono già, alcuni dei quali descritti proprio in questo convegno (si pensi agli innumerevoli progetti promossi da Ca' Foscari negli ultimi decenni),

14 Alla Biblioteca Nazionale Centrale di Firenze un Gruppo di lavoro (a cui partecipa l'Accademia della Crusca) sta approfondendo i criteri da impiegare nella soggettazione di opere antiche, con l'obiettivo di pubblicare delle Linee guida. Un'interessante proposta di impiego del Nuovo soggettario per l'indicizzazione semantica di risorse antiche, peraltro di natura documentaria e archivistica, si trova in Grillo 2015.

15 URL <http://www.lex.unict.it/cir/#bianet> (2017-10-26).

oppure in questa stessa Tavola rotonda (ad esempio, la Biblioteca digitale di testi latini tardoantichi, presentata da Maurizio Lana), ma che non sempre hanno ancora avuto occasione di interfacciarsi.

In questo passaggio a una collaborazione che coinvolga maggiormente il mondo della ricerca e grazie alla quale mettere a disposizione strumenti più rigorosi sul piano scientifico, più efficaci e aperti, le Biblioteche nazionali, se volessero, potrebbero davvero giocare un ruolo importante, favorendo l'avvio di progetti comuni, nell'ambito dei quali 'mescolare' competenze e facendo lavorare insieme categorie professionali diverse, un'altra chance affascinante, ancorché troppo sottovalutata, offertaci dalle biblioteche digitali.¹⁶

¹⁶ I dati riportati in questo articolo sono riferiti e aggiornati al mese di agosto 2015.

Bibliografia

- Balíková Marie et al. (2015). «Facilitating Access to Cultural Heritage Content in Czechia. National Authority Files and Interim Project» [online]. URL <http://library.ifla.org/1135/1/148-balikova-en.pdf> (2017-10-25).
- Bergamin, Giovanni; Lucarelli, Anna (2013). «Il Nuovo soggettario come servizio per il mondo dei linked data» [online]. *JLIS.it. Italian Journal of Library, Archives, and Information Science*, 4(1). URL <http://leo.cilea.it/index.php/jlis/article/view/5474/7904> (2017-10-25).
- Biagetti, Maria Teresa (2014). «Sviluppi e trasformazioni delle biblioteche digitali. Dai repositories di testi alle semantic digital libraries» [online]. *AIBstudi*, 54(1), 11-34. URL <http://aibstudi.aib.it/article/view/9955> (2017-10-25).
- Biblioteca Nazionale Centrale di Firenze (2007). *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto. Prototipo del Thesaurus*. Milano: Editrice Bibliografica.
- Grillo, Manuela (2015). *Indicizzazione semantica di bandi, manifesti e fogli volanti*. Sassari: Documenta.
- Lucarelli, Anna (2014). «Nuove scommesse della BNCf. Wikipediani in residence, Wikisource e altro ancora» [online]. *Digitalia web. Rivista del digitale nei beni culturali*, 2, 100-6. URL <http://digitalia.sbn.it/article/viewFile/1292/849> (2017-10-25).
- Lucarelli, Anna; Viti, Elisabetta (2015). «Florence-Washington Round Trip. Ways and Intersections Between Semantic Indexing Tools in Different Languages». *Cataloging & Classification Quarterly*, 53(3-4), 1-16.
- Tammaro, Anna Maria (2014). «Convergenza di biblioteche, archivi e musei. Le iniziative di IFLA» [online]. *AIBstudi*, 54(1), 115-20. URL <http://aibstudi.aib.it/article/view/10004> (2017-10-25).

Alcuni problemi teorici e pratici posti dalle *digital libraries*

Raul Mordenti

(Università degli Studi di Roma "Tor Vergata", Italia)

Abstract In this article we will point out the substantial difference between 'digital edition' and 'informatic edition': the first one (DH = Digital Humanities) is the simple digitalisation of a manuscript or a printed text; the second one (HC = Humanities Computing) is an edition created by taking into account the potential of the computer itself and with the intention of exploiting the texts completely. This implies that the informatic edition is designed and built in accordance with the basic laws that allow us to give instructions to a computer (completeness, expediency, univocity, no contradiction, no redundancy, etc.). The key word of an informatic edition is encoding, which is the basis for the scientificity of the edition. Moreover, the possibility of using the computer and its potential to enjoy the text depend on it. Finally, we will deal with the different ways of disseminating the texts, noting some radical differences between the traditional editorial circuit, which is typical of the printed text, and the new one established by computers and new technologies (first of all, the indefensibility of copyright).

Sommario 1 Per una definizione più rigorosa di DH e HC. – 2 La nuova biblioteca digitale: caratteristiche, problemi e necessità.

Keywords Digital humanities. Digital library. Text. Copyleft. Humanities computing.

1 Per una definizione più rigorosa di DH e HC

Vorrei sollevare due soli problemi, che mi sembra siano già emersi – sia pure in forma implicita e indiretta – in molti interventi: 1) il problema di una definizione teorica più precisa della disciplina che (ne siamo noi coscienti oppure no) stiamo fondando, e 2) il problema del rapporto della nostra collettiva biblioteca digitale con quella cartacea e guttemberghiana di cui essa prenderà il posto.

Non sembri trascurabile lo sforzo di una precisazione terminologica dato che essa comporta necessariamente anche una importante precisazione concettuale. La distinzione fondamentale che si propone è quella fra HC (Humanities Computing) e DH (Digital Humanities).

Dico subito che la necessità di tradurre in italiano queste due espressioni non ci aiuta, giacché mi sembra che fra noi manchi a tutt'oggi una traduzione condivisa di queste espressioni inglesi (una traduzione potreb-

be forse essere, rispettivamente: 'Informatica Umanistica' e 'Umanistica Digitale', ma si aspettano proposte migliori); così come non ci aiuta il fatto che la formula DH (forse semplicemente per la sua comodità) sembra essere entrata ormai nell'uso comune, fino a partecipare di molti acronimi di importanti istituzioni del nostro settore, sia di ambito europeo, come EADH (The European Association for Digital Humanities), che tuttavia ha mutato il nome originale (risalente al 1973) di ALLC (Association for Literary and Linguistic Computing), sia anche di ambito mondiale, come ADHO (Alliance of Digital Humanities Organizations); ma mi piace notare che la nostra italiana AIUCD (Associazione per l'Informatica Umanistica e la Cultura Digitale), costituita a Firenze nel 2011, ha scelto il suo nome, dopo un lungo dibattito, proprio per unire i due ambiti, cioè - se ho ben inteso - quello più specifico e proprio dell'Informatica Umanistica e quello più vasto della Cultura Digitale.

Il punto è che, come ci ha insegnato Tito Orlandi, esiste effettivamente una differenza sostanziale fra il 'Digitale' (da cui DH, Digital Humanities) e il 'Computazionale' (da cui HC, Humanities Computing). E proprio il nostro campo della filologia e delle edizioni, si presta a chiarire, come meglio non si potrebbe, di quale differenza si tratti. Un'edizione digitale, o digitalizzata, è la semplice trasformazione di un oggetto analogico (nel nostro caso un manoscritto o una stampa o altro) in un oggetto digitale, e questo - almeno a partire dall'invenzione degli scanner - può anche avvenire senza alcun intervento critico da parte dello studioso, più o meno come avveniva per fare una fotocopia.

Un'edizione computazionale (che io chiamerei semplicemente informatica) è invece un'edizione che si svolge a partire dalle potenzialità euristiche della computazione (cioè della macchina di Turing) con l'intenzione di sfruttarle pienamente per migliorare la conoscibilità del testo. Ciò comporta due diversità fondamentali fra HC e DH (che non significano affatto - vorrei essere del tutto chiaro su questo punto - una qualche gerarchia fra HC e DH).

Anzitutto solo nel primo caso (HC), e non nel secondo (DH), l'uso dell'informatica si pone un obiettivo propriamente euristico, cioè intende conseguire un significativo plusvalore conoscitivo, che può consistere nel produrre edizioni critiche migliori o nel consentire operazioni analitiche impossibili senza l'informatica, nel definire ontologie adeguate ai nostri campi di ricerca o nel supportare originali analisi semi-automatiche, ecc. È la computazione che fa del computer, come è stato detto, un 'medium cognitivo'. Sembra invece caratterizzare DH in incremento (voglio ripeterlo: altrettanto utile e prezioso) della comodità di accesso ai dati e della loro diffusione, come ad esempio la produzione delle grandi e ormai indispensabili banche dati testuali.

In secondo luogo solo nel primo caso (HC), e non nel secondo (DH), è implicata, e anzi diventa centrale, la procedura della ricerca, più precisa-

mente la riutilizzazione e la ridefinizione delle procedure caratteristiche delle nostre discipline umanistiche che sono chiamate a operare nel nuovo ambiente informatico. Personalmente credo che proprio questo sia il punto più prezioso (e – se posso dirlo – anche entusiasmante) del nostro attuale lavoro, quello in cui le nostre discipline umanistiche e il loro secolare corredo di procedure, venendo messe alla prova nel nuovo ambiente segnato dall'informatica, ritrovano, ma anche modificano e affinano, se stesse.

Penso anche che consista esattamente in questo punto un apporto possibile degli umanisti, che non si limitano dunque più a 'ricevere' e usare qualcosa che proviene dagli informatici ma possono a loro volta fornire qualcosa di inedito, per ipotesi utile allo sforzo comune di incrementare la conoscenza.

Limitandoci al nostro campo che riguarda la filologia e i testi, si potrebbe dire che la Digitalizzazione lavora sull'oggetto della ricerca (ora reso in forma digitale e non più analogica) e si arresta invece di fronte alle porte del testo in quanto tale, cioè in quanto messaggio storico e semantico (accade questo, ad esempio, quando vengono fornite in formato digitale le immagini dei manoscritti), mentre la Computazione intende proprio attraversare quella soglia fatale per incrementare, attraverso l'uso della macchina informatica, le nostre capacità di lettura dell'oggetto testuale.

Estremizzando forse il nostro ragionamento, credo che potrebbero esistere forme di Digitalizzazione che non prevedono a rigore Computazione, mentre la Computazione presuppone sempre la Digitalizzazione. Più precisamente: la Digitalizzazione si pone all'inizio come premessa necessaria della Computazione (dato che la macchina informatica lavora solo su dati digitali) e poi necessariamente anche alla fine (dato che intendiamo comunque proporre in formato digitale i dati da noi elaborati). Ma proprio questa successione lineare, o intreccio, Digitalizzazione-Computazione-Digitalizzazione, in cui però – si noti bene – nessuno dei due momenti è compreso logicamente all'interno dell'altro, ci aiuta a capire che stiamo effettivamente parlando di due cose diverse. Che dunque non vanno confuse, né dal punto di vista logico né dal punto di vista terminologico.

Vorrei aggiungere che tale differenziazione (per me sacrosanta quanto per lo più rifiutata) fra Digitalizzazione e Computazione (o meglio ancora fra ambiente digitale e ambiente computazionale), (a) non coincide con la differenza fra singole procedure, in quanto nella Digitalizzazione può anche essere compresa una parte computazionale, e nella Computazione – come abbiamo appena detto – è sempre compresa una parte digitale; (b) essa dipende dunque dalla considerazione soggettiva dei creatori e degli utenti piuttosto che dalla oggettiva situazione della realtà.

Cercando di essere un po' più esplicito, e rimanendo nell'ambito filologico e linguistico, un primo esempio è dato proprio dalla Digitalizzazione di un testo. Essa si può considerare come la sostituzione di un simbolo con un altro simbolo; ma si dimentica spesso la lezione di Peirce, che un simbolo è

tale solo in presenza di un interpretante: esso cioè mette in moto una serie di competenze senza le quali il simbolo non è più tale ma un'inerte forma grafica che sottoposta a procedimenti automatici darà risultati per lo meno equivoci. Per questo la parola-chiave imprescindibile è, nella nostra procedura, codifica. Questo vale per la singola lettera, le cui diverse forme possono avere diversi valori (significati) e dunque devono essere rappresentate da simboli digitali diversi, quanto per le 'parole', che sono simboli utilizzabili in procedure automatiche soltanto se digitalizzate non solo in base alle loro componenti alfabetiche, ma distinguendo le caratteristiche che esse portano con sé nel testo di cui fanno parte. E vale anche per le altre unità di rango superiore, le frasi, le sezioni (paragrafi, capitoli, ecc.), le pagine (valore dell'impaginazione) e così via. Tutto questo non vale se si vuole fornire all'utente soltanto una rappresentazione brutale del testo nel suo senso più banale; e naturalmente anche questo può essere utile per molti usi, ma non per una procedura computazionale, sia pure una semplice estrazione di elementi (peggio se con conseguenti statistiche), perché verranno giudicati equivalenti fenomeni fra sé simili per certi aspetti, ma diversi per altri. In ambito generale umanistico questo momento analitico che precede logicamente la Computazione assume dunque un'importanza fondamentale, perché richiede un'attenzione e una consapevolezza molto speciali, mentre per noi il momento computazionale vero e proprio è relativamente poco complesso. Ma anche qui, sarebbe opportuno valutare le trovate che vanno di moda in confronto alla tradizione di studi, millenaria, che ci appartiene. La stemmatica genealogistica, le cosiddette 'sentimental analysis' e 'network science', siano pure procedimenti computazionali di pieno diritto, ma mi sembrano scorciatoie puerili per pretendere di superare metodologie tradizionali di gran lunga più serie.

2 La nuova biblioteca digitale: caratteristiche, problemi e necessità

Il secondo problema che vorrei sollevare (altrettanto rapidamente e, ahimé, altrettanto sommariamente del primo) riguarda il rapporto, e la differenza, fra la biblioteca digitale che intendiamo costruire insieme e quella gutenberghiana in cui abbiamo finora operato.

Non mi sembra interessante partecipare al gioco mediatico in ordine alla pretesa 'morte del libro' cartaceo: Umberto Eco ha detto sensatamente che il libro è come il cucchiaino, cioè è un oggetto talmente comodo e funzionale che sopravviverà, magari in posizione marginale, anche nel nuovo assetto (i dati recenti sulla precoce crisi di vendita degli e-book sembrano dargli ragione) ed è comunque vero che l'invenzione dell'automobile non ha comportato automaticamente la fine dei cavalli.

Il problema su cui vorrei richiamare l'attenzione è un altro, e riguarda la crisi (e io penso: la fine) del circuito editoriale che si era costruito a partire dal libro di Gutenberg. Dice un adagio caro agli editori che l'editore è colui che non scrive il libro, non lo stampa, non lo distribuisce, non lo vende, non lo recensisce, ma... fa tutto il resto. Se cerchiamo di analizzare questo 'resto', assolutamente decisivo nell'assetto editoriale gutemberghiano, potremmo dire che l'editore (parlo naturalmente dell'editore degno di questo nome, dell'editore vero e proprio): a) progettava/sceglieva/ programmava il testo, ad esempio inventandosi le collane (come il nostro Manuzio) o addirittura commissionando la scrittura di determinati libri agli autori; b) 'curava' la pubblicazione, intervenendo direttamente sul testo, attraverso l'attività redazionale (e non c'è bisogno di ricordare qui che ruolo abbiano svolto nella cultura italiana i redattori, dal Bembo delle edizioni 'aldine' fino a Calvino o Bassani); c) garantiva la qualità tipografica, l'accuratezza, l'attendibilità, e a volte anche una peculiare bellezza dell'oggetto libro; d) assicurava una mediazione vitale fra il libro e il mercato dei lettori, promuovendo il libro, curando le presentazioni e/o le recensioni, ecc. e, soprattutto, garantendo in un certo senso la qualità del libro attraverso il prestigio del proprio marchio editoriale; e) garantiva infine la reperibilità del libro, e dunque la sua durata, attraverso il proprio magazzino ma soprattutto attraverso il proprio catalogo, che rappresentava la vera gloria di un editore di qualità.

Era l'insieme di queste attività, come si vede assolutamente cruciali, ciò che giustificava e legittimava il profitto dell'editore.

Ora, se esaminiamo il nuovo circuito editoriale che si sta determinato a partire dai prodotti testuali informatizzati, ci rendiamo conto che nessuna di queste funzioni viene più svolta dagli editori (le eccezioni o le tenaci sopravvivenze non interessano questo nostro sommario discorso): a) quasi nessun editore può più progettare, investendo per il medio-lungo periodo, perché le leggi del mercato non glielo consentono; b) le redazioni, il vero centro produttivo delle case editrici di una volta, sono state le prime a essere sciolte (o esternalizzate, che è lo stesso) per risparmiare sui costi; c) della qualità tipografica è meglio non parlare, e non sono poche le case editrici che pretendono dagli autori il libro già impaginato, in PDF, completo di titoli correnti e indici, limitandosi a stamparlo e a fornirlo di ISBN; d) non esiste praticamente più alcuna mediazione di tipo culturale fra libro e lettori (evidentemente l'eventuale pubblicità è tutt'altra cosa), anzi sono sempre più frequenti i casi in cui il libro è pagato, direttamente o indirettamente, dal suo stesso autore, magari usando per questo i sempre più esigui fondi destinati alla ricerca; e) infine - forse questa è per noi la circostanza più negativa - il catalogo non esiste quasi più, la durata del libro è una chimera, i libri durano lo spazio di un mattino e il macero li attende; questa rapida irrimediabilità riguarda addirittura l'opera omnia di alcuni classici (cito per tutti la fondamentale serie delle *Opere complete*

di Boccaccio, curata da Branca per Mondadori, un'edizione che doveva essere per sua natura *für ewig* e che risulta invece assolutamente introvabile già a pochi anni dalla sua pubblicazione).

Tutto ciò, e in particolare il per noi cruciale punto e), cioè la mancata garanzia della durata, risulta enfatizzato in sommo grado per le opere in formato digitale; anzi neanche l'acquisto garantisce il possesso nel tempo del libro digitale, se è vero che spesso se ne compra solo il diritto alla lettura che, in quanto tale, non può essere neanche trasmesso ai nostri eredi.

Per non parlare del problema decisivo rappresentato dalla obsolescenza dei testi digitali, che riguarda non tanto i loro supporti quanto i programmi e le macchine necessari per leggerli. Faccio notare che la biblioteca di pubblica lettura nel vecchio assetto svolgeva un ruolo correttivo del mercato (appunto perché garantiva la lettura gratuita e la conservazione nel tempo del libro), ma essa non è sostituita nel nuovo assetto da alcuna istituzione paragonabile. Già ora, negli scaffali di ciascuno di noi, esistono vecchi floppy disk che risultano inaccessibili ai nuovi computer, determinando una perdita secca di informazioni testuali che nel tempo potrebbe assumere dimensioni catastrofiche. Nel mondo dell'informatica e dei suoi programmi sempre mutanti, è come se si verificasse ogni pochi anni una modificazione della tecnologia testuale paragonabile a quella che segnò il passaggio dal rotolo al volume o dal manoscritto alla stampa. E ogni volta - come sappiamo bene - queste modificazioni tecnologiche della testualità comportarono, assieme alla necessità del 'riversamento' dei testi da una tecnologia all'altra, anche perdite testuali enormi e irreparabili. Ma soprattutto, nel nuovo circuito editoriale segnato dall'informatica, l'editore dove sta? In cosa consiste la sua utilità? Cosa legittima la sua pretesa di trarre profitto dai libri?

Credo, in conclusione, che sia necessario immaginare e implementare al più presto un originale sistema di circolazione e fruizione del libro elettronico che tenga conto delle sue novità e delle sue specifiche caratteristiche e potenzialità, insomma un circuito che non imiti più il vecchio assetto gutenberghiano ma che si svolga *iuxta propria principia*, cioè che sia in grado di garantire di fatto ciò che l'informatica (in questo caso la rete) è per sua natura in grado di consentire, ma dunque in qualche modo pretende che sia realizzato

Mi sembra che ciò significhi anzitutto garantire l'accesso universale e dunque in primo luogo la gratuità, considerando la cultura per quello che effettivamente è, cioè come un bene comune. D'altra parte la cultura è forse l'unico bene che si incrementa con l'uso collettivo e per il quale non vale affatto il principio di esclusività (se una cosa è tua non è più mia, e viceversa), il principio che valeva per i campi e per il bestiame, anzi in questo caso il fatto che molti fruiscano della mia cultura mi arricchisce e non mi impoverisce, esattamente come mi arricchisce fruire della cultura altrui. Questo chiama in causa la mano pubblica, e la necessità di specifiche iniziative e di adeguati finanziamenti pubblici.

Le Università pubbliche, che già oggi ospitano, in modo generoso ma non sistematico e non sostenuto, le nostre biblioteche digitali e i prodotti digitali delle nostre ricerche, devono essere il luogo di questa nuova biblioteca di Alessandria informatica, universale, diffusa e accessibile a tutti.

Mi rendo conto che si tratta di un'affermazione controcorrente in un'epoca che pretende di fare del mercato e delle sue pretese la nuova Legge universale e totalitaria. Ma se è vero - come è stato detto da chi se ne intendeva - che la guerra è una cosa troppo seria per lasciarla fare ai generali, credo che sia ancora più vero che la cultura digitale è una cosa troppo seria per lasciarla nelle mani della spontaneità capitalistica, del mercato e degli interessi delle grandi imprese multinazionali dell'informatica.

Nota conclusiva

Raffaele Perrelli

(Università della Calabria, Italia)

Un aspetto degno, a mio avviso, di una pausa di riflessione alla fine di queste giornate di studio suonerà un po' datato, ma mantiene intatta la sua capacità di porre questioni più importanti delle risposte che si attendono, di rappresentare, cioè, un ambito di valori culturali cui far riferimento. Mi riferisco alla dimensione 'storica' dello studio dei testi antichi ('storicistica' si sarebbe detto, più ideologicamente, alcuni anni fa): una dimensione che comprende la contestualizzazione in senso stretto, necessaria a evitare ogni travisamento del senso, e in maniera più generale, a evitare anacronismi e forzature nella lettura dell'antico. Penso alla libera circolazione di testi nella rete al di fuori delle cornici 'controllate' delle edizioni digitali, di cui oggi si è parlato, alla presenza di testi antichi privi di alcun controllo nelle tante circostanze di scrittura e lettura (blog, social, ecc.) che la rete offre.

Può sembrare un passo indietro rispetto a molte delle posizioni e dei riferimenti emersi in queste giornate ma non credo sia così. La diffusione delle *Digital Humanities* ha comportato una circolazione libera dei testi nella rete. Definisco ora 'libera'. Odi oraziane, carmi catulliani, lacerti virgiliani sono citati in funzione del significato 'letterale' della citazione. Questo significato letterale tende all'universalizzazione dei significati e, pertanto, anche alla smaterializzazione dei contesti. Questo aspetto mi sembra il più pericoloso.

Faccio due esempi: come è noto, Baudelaire, in un suo famoso componimento (*Le cygne*), riprende l'episodio di Andromaca nel terzo libro dell'*Eneide*. Sorprendono il rigore di quella ripresa, l'attenzione ai versi virgiliani seguiti da vicino. Nel *Passaggio di Enea*, Caproni ricorda la convulsa fuga di Enea nel finale del secondo libro, con Anchise sulle spalle e Ascanio accanto. Tanto la ripresa di Baudelaire è vicina al testo latino e segue la caduta della gloria di Andromaca, il suo precipizio matrimoniale da Ettore fino a Eleno, quanto quella di Caproni è evocativa, vaga, generica. Non sto giudicando o confrontando due capitoli della fortuna. Ma solo ricordando che i testi classici hanno sempre goduto di uno spettro molto ampio di riuso. I commenti tradizionali, tuttavia, hanno svolto proprio questa funzione, vale a dire colmare uno iato di significati e contesti o ricordare la vicinanza della ripresa all'originale. La libera circolazione

delle merci verbali nel mercato della rete rende pressoché impossibile, oltre che sgradito (ma questo è sempre stato vero), questo tipo di intervento. Bisogna tener conto che la nuova vita delle 'parole dell'antico' sarà forse più fortunata di quella del recente passato, ma si tratta di parole del passato e senza passato al tempo stesso, figlie di una cogiacenza testuale che mette ai margini la filologia e la storia.

Il libro costituisce l'esito di un Convegno internazionale di studi tenutosi all'Università Ca' Foscari Venezia nel 2014, e si propone quale momento di riflessione sopra le opere di digitalizzazione e ricerca condotte a livello accademico su materiali e testi inerenti le Scienze dell'Antichità. I contributi in esso raccolti rispondono all'esigenza di porre a confronto esperienze maturate in ambiti disciplinari apparentemente distanti fra loro (dall'archeologia all'epigrafia, dalle letterature antiche e premoderne al diritto romano) ma accomunate da punti di contatto significativi. Lo scopo è di individuare nella collaborazione interdisciplinare nuove linee di ricerca, capaci di estendere le potenzialità degli attuali sistemi automatici di indagine.



Università
Ca'Foscari
Venezia