

Tito Orlandi

UN ULTIMO BILANCIO DELL'INFORMATICA UMANISTICA

È opportuno prima di tutto che spieghi il significato del titolo, che anche vorrebbe essere il significato generale di questo intervento. Ultimo bilancio: ultimo è da intendersi in senso soggettivo, in quanto non ne farò altri, anzi intendo da questo momento in poi agire nell'ambito dell'informatica umanistica solo per la parte che mi riguarda specificamente e concretamente, rinunciando a propagandare quei principi che pure mi sembrano più che mai validi, ma che lasciano indifferente, mi si lasci dire colpevolmente indifferente, la maggior parte del mondo accademico o comunque dei responsabili delle strutture accademiche, e che sono presto riassunti: la visione unitaria dell'Informatica Umanistica, il rigore nell'uso dei mezzi digitali per la ricerca, la scelta di un ambiente operativo "aperto" (open, dunque unix e non windows), il riconoscimento universitario della disciplina. Ma bilancio da intendersi in senso oggettivo, in quanto cercherò di proporre una serie di giudizi oggettivamente validi, e sperabilmente utili nei diversi settori considerati, anche nell'intento pragmatico del conoscere per deliberare.¹

A questo scopo intendo tracciare un panorama storico dello sviluppo dell'Informatica Umanistica, che mi sembra si possa utilmente dividere in quattro periodi: quello dei precursori, quello delle applicazioni pionieristiche, quello degli esperimenti e delle ricerche, quello del trionfo della tecnologia.

Periodo dei precursori (fino al 1960). Per lungo tempo ci si è addirittura meravigliati che i computer potessero trovare un qualsiasi utilizzo nell'ambito delle discipline umanistiche. È dunque naturale che in questo periodo troviamo due soli progetti da prendere in considerazione, Uno è quello volto ad automatizzare il procedimento della traduzione, ed è stato proposto da William Weavers nel 1947;² l'altro quello di produrre l'indice delle opere di s. Tommaso, ed è stato proposto da Roberto Busa nel 1946, per poi essere ampliato per produrre le concordanze.³

Di là da una certa casualità nella scelta del tipo di applicazione, come è ovvio in ogni fase iniziale, ci troviamo di fronte a due interessanti caratteristiche, che si ritroveranno, insieme ad altre, nello sviluppo successivo dell'Informatica Umanistica.

La prima consiste nel fatto che che l'accostamento ai computer è guidato da problemi sorti prima della nascita del computer, la cui soluzione si è pensato potesse essere facilitata dall'uso della macchina: l'esigenza di tradurre, e in altro campo indicizzare, una vasta mole di testi. Questo prelude alla classica illusione che accompagnerà il cammino dell'Informatica Umanistica, per la quale l'uso del computer dovrebbe necessariamente e semplicemente rendere più agevole e rapido svolgere alcuni dei compiti, lasciando inalterate tutte le altre condizioni. È invece ormai assodato che il non tener conto del fatto che le applicazioni informatiche determinano il mutamento di molte delle condizioni iniziali è letale per le ricerche in questo campo.

La seconda caratteristica è la distinzione fra il perseguimento di procedure propriamente computazionali e l'utilizzazione di procedure di puro ordinamento e calcolo numerico, che fra l'altro si richiamano a due strade diverse, e solo in un secondo tempo convergenti, nella nascita degli stessi computer. Da un lato il progetto di un calcolatore elettronico digitale binario (ENIAC, Moore School, Pennsylvania),⁴ dall'altro quello di un ordinatore e contatore di schede perforate (IBM).⁵

Per quanto basata su presupposti linguistici erronei, l'idea della traduzione automatica nasceva dalla presa di coscienza del calcolatore elettronico digitale come esecutore di procedure computazionali, ed era molto più produttiva ed interessante, dal punto di vista dei metodi applicativi, che non la semplice automazione lessicografica, nata dall'ordinatore a schede, anche se il prodigioso avanzamento tecnologico finirà col concentrare l'attenzione su questo secondo aspetto rispetto all'altro.

Periodo delle applicazioni pionieristiche (1960-1970). I computer cessano di essere oggetto di meraviglia, almeno presso alcuni studiosi illuminati, e intraprendenti umanisti in maggior numero cercano di sperimentare le capacità delle macchine. Indicizzazione, ordinamento, e statistica lessicale, nelle loro varie tipologie, sia pure sotto nomi variamente fantasiosi, sono il campo d'azione principale: in Italia con Alinei e Avalle; in ambiente anglosassone con Wisbey, Ellison, Morton, e Wake; in Francia con Quemada.

È da sottolineare tuttavia il nascere dell'idea dell'applicazione in ambito ecdotico,⁶ con le proposte di Froger, e di Maretti e Zarri, che seguono il filone più propriamente computazionale, anche se con ingenuità metodologiche che del resto si ritrovano fino ad oggi.

Altrettanto interessante l'applicazione in ambito archeologico, con l'apparizione del grande Jean-Claude Gardin, che comincia a cercare linguaggi per l'analisi di fenomeni non linguistici.

Periodo degli esperimenti e delle ricerche (1970-1990). È questo, secondo me, il periodo più vivace ed interessante

della storia che stiamo seguendo. Per quanto concerne i fenomeni testuali, l'analisi lessicale si amplia all'analisi delle strutture linguistiche, con la nascita della linguistica computazionale. Ricordo Zampolli, Berni Canani, Geoffrey Sampson, Susan Hockey. Quest'ultima deve essere ricordata, insieme con l'ambiente di Oxford, anche per software pionieristici come il linguaggio Snobol e l'OCF. L'attività in questo campo fu molto vasta.

Gardin e Borillo sviluppano la visione dell'Informatica Umanistica come un nuovo linguaggio rigoroso, formalizzato, adatto a descrivere fenomeni archeologici, storici, e letterari, e in maniera meno sofisticata qualcosa di simile sperimenta a Pisa Paola Barocchi per la storia dell'arte.

Manfred Thaller sviluppa i temi della codifica e delle basi di dati in ambito storiografico, anticipando con il suo Kleio i temi della modellizzazione dei fenomeni.

In ambito di storia dell'arte appaiono gli esperimenti arditissimi di Corrado Maltese e di Letizia Zanobini.

Un settore molto delicato è quello della statistica (lessicale o fattuale, in storia ed archeologia: da ricordare Amilcare Bietti), perché non è facile distinguere fra un uso puramente matematico del computer ed uno slancio verso l'ideazione di algoritmi, di metodi di valutazione dei dati analizzati.

La maggior parte di queste ricerche fu purtroppo destinata a rimanere, almeno fino ad oggi, senza conseguenze di ricerca strutturata e riconosciuta, a causa soprattutto della refrattarietà e dell'ignoranza informatica del mondo accademico, non soltanto in Italia, ma in tutto l'ambiente internazionale. Gardin mi confessava anni fa tutto il suo sconforto; Thaller, guadagnato un posto di ordinario all'Università di Köln, è ridotto a curare pubblicazioni digitali di fotografie di codici antichi. Ricorderò che le due sole annate pubblicate dello *Humanities Computing Yearbook* (1988 e 1989-90) cercavano di dare contezza di tutto questo fermento, e aggiungerò sommessamente che anche il mio libro, intitolato appunto *Informatica Umanistica*, pubblicato nel 1990, cercava di rispecchiarlo, ma è rimasto impietosamente ignorato.

Periodo del trionfo della tecnologia (1990 ad oggi). Gli incredibili avanzamenti della tecnologia hanno messo fra le mani degli studiosi (in senso non figurato, se pensiamo ai computer portatili) strumenti di una potenza ignorata dai più grandi main frame degli anni precedenti, e sono oltretutto già dotati di software più o meno incorporati, che hanno fatto dimenticare i semplici ma essenziali principi su cui il loro funzionamento si basa, ed hanno limitato il loro apprezzamento ai programmi che essi offrono, eclatanti dal punto di vista dell'utente finale, col risultato di focalizzare l'attenzione sul risultato visibile, e non sulle procedure sperimentali che dovrebbero essere l'oggetto della ricerca scientifica anche per le discipline umanistiche.

Si parla dunque di tecnologia anche per quanto riguarda il software (cosa che prima non succedeva), perché esso appare inalterabile quanto la macchina a cui è legato, ed in questo senso parliamo con rincrescimento di trionfo della tecnologia, e crediamo si possano classificare in tre grandi gruppi le realizzazioni che prendiamo in considerazione: strumenti, prodotti, e istituzioni. Su ognuno di essi dovrei soffermarmi a lungo, per giustificare i giudizi che esprimerò, ma in questa sede mi si perdonerà se mi limito a proporli come aforismi, lasciando la discussione a momenti propizi e a chi vi sia particolarmente interessato.

(a) Strumenti. Sono nati per aiutare e normalizzare le principali operazioni con cui si confronta l'Informatica Umanistica. Prima di tutto la codifica digitale e il conseguente markup (annotazione), per i quali sono da menzionare insieme SGML,⁷ XML,⁸ e TEI.⁹ Sono questi dei linguaggi descrittivi la cui concezione è importantissima, l'uso necessario, ma di per sé non producono nulla. Essi rimandano all'analisi dei dati, che in genere non viene fornita da coloro che li utilizzano, e sembra essere l'ultima delle loro preoccupazioni.

Il protocollo TEI vorrebbe coprire tutte le possibilità dei fenomeni testuali, e per questo è troppo complesso, oltre naturalmente a non poterle esaurire, direi, in linea di principio. Nasce da ciò il problema delle gerarchie multiple, cioè delle strutture testuali fra sé incompatibili, che non ha nulla di particolarmente complicato, purché si abbandoni la pretesa tipica della TEI di essere onnicomprensiva. Essa del resto è nata soprattutto per la descrizione di testi che vengono prodotti, e non di quelli storicamente esistenti, e a questo compito dovrebbe essere limitata.

Aggiungeremo a questo proposito anche l'operazione *Unicode* (volta alla standardizzazione delle polizze di caratteri), che non viene capita dagli utilizzatori meno avvertiti. Essa ha un valore puramente pratico, di rapporto fra utilizzatore e software di scrittura, mentre dal punto di vista della digitalizzazione scientifica del testo non ha alcuna rilevanza.

In sostanza bisogna denunciare una disattenzione verso i problemi di metodo riguardanti la codifica dei testi, o peggio ancora il loro travisamento, dovuto alla mancanza di una seria riflessione sul concetto di testo.

(b) In secondo luogo ricordiamo l'affermarsi del protocollo world wide web e del linguaggio ipertestuale HTML.¹⁰ È superfluo parlare dell'importanza epocale del fenomeno, su cui si versano fiumi d'inchiostro, ma bisogna dire

chiaramente che dal punto di vista della ricerca il suo significato dipende dalla qualità dei contenuti, e dunque di per sé non va nemmeno preso in considerazione.¹¹

In terzo luogo ricordiamo l'affermarsi della metodologia relazionale nella ideazione, costruzione, e gestione delle banche-dati. Questa tecnologia ha segnato un progresso importantissimo nella concezione e nella gestione delle banche-dati, ma ancora una volta è di solito male inteso da affrettati utilizzatori. Infatti non si comprende che essa è un modo di analizzare la realtà secondo la teoria computazionale, e questa analisi che (cattiva o buona) è alla base del funzionamento delle banche dati non è quasi mai dichiarata.

A questo proposito menzioneremo i GIS, una accattivante e complessa tecnologia per trattare la documentazione cartografica e archeologica.¹² Anche di essi si parla a sproposito come fossero dei risolutori di problemi. In realtà sono la mescolanza, permessa dal sistema digitale, di banche dati, riproduzione visiva, e anche testo, e dunque rimandano alla correttezza delle parti che li compongono.

(c) Prodotti. Sono in certo senso i risultati ottenuti mediante gli strumenti menzionati sopra. Ne ricorderemo due tipi che in questo periodo hanno avuto particolare estensione:

(a) Corpora testuali. Riguardano da un lato testi tratti dal linguaggio corrente, da giornali o da voce registrata. Qui si notano le novità più interessanti, in quanto hanno suscitato una polemica molto produttiva circa il loro valore per le ricerche di linguistica. Tuttavia per la loro codifica non pare utilizzato correttamente lo strumento essenziale dell'alfabeto fonetico, che potrebbe fornire importanti indicazioni per le ricerche sui rapporti fra linguaggio e scrittura.

Sono poi stati prodotte versioni digitali di testi letterari o comparabili, come l'Oxford Text Archive,¹³ il Thesaurus Linguae Graecae,¹⁴ la Letteratura Italiana Zanichelli,¹⁵ etc. In tutti questi casi, se sul piano pratico dell'utilizzazione, diciamo così, brutale, gli esiti si sono rivelati notevoli, sul piano della correttezza testuale e filologica, ed ancor più della teoria delle trasformazioni del testo (manoscritto, stampato, digitalizzato) la conseguenza dei metodi adottati è stata soprattutto quella di uccidere sul nascere le riflessioni teoriche avanzate. Preferisco qui non reiterare polemiche svolte altrove. Ma bisognerà pure distinguere fra un'attività di ricerca meditata ed una di pura organizzazione a fini pratici.

(b) Archivi enciclopedici, del tipo, per intendersi, della Wikipedia. In questo, come in altri casi menzionati sopra, la valutazione scientifica riguarda il contenuto, dunque altre discipline tradizionali, e l'Informatica Umanistica ne è interessata soltanto in modo collaterale. Del tutto a sé va considerato piuttosto il *Principia Cybernetica Project*,¹⁶ tentativo di interesse eccezionale di esporre mediante un sistema cosiddetto ipertestuale il mondo concettuale della cibernetica, che è essenziale per le metodologie computazionali. A quanto mi consta esso rimane ignorato, comunque non discusso, nell'ambiente dell'informatica umanistica, per il quale esso sarebbe invece fondamentale.

Parallelamente al diffondersi della tecnologia, nel senso spiegato sopra, si sono moltiplicate le istituzioni, intendendo per tali i luoghi, reali o virtuali, in cui si promuovono le attività di Informatica Umanistica. Centri operativi e di consulenza sono sorti numerosi presso numerose Università. In Italia il primo è stato il CISADU (Roma Sapienza), e oggi sono in funzione analoghi centri a Pisa (CISIAU),¹⁷ Torino (CISI),¹⁸ Milano (SIPU).¹⁹

Per l'estero non è il caso di fare elenchi; diremo soltanto che in questo settore particolarmente vivace è il Regno Unito, e notevoli realizzazioni si rilevano negli Stati Uniti, in Canada, e in Olanda.

L'insegnamento dell'Informatica Umanistica è apparso relativamente tardi nel panorama delle attività ad essa relative, potendosi far risalire a non prima del 1985 circa. Per la verità furono proprio quei primi tempi, fino al 1995, che videro le iniziative accademiche più interessanti, che io stesso ho registrato in una inchiesta condotta per il progetto europeo AcoHum (Advanced Computing in the Humanities),²⁰ pubblicata nel 1999.²¹

In seguito, perduto l'interesse per la ricerca metodologica di base, l'insegnamento si è ristretto per lo più alla cosiddetta alfabetizzazione informatica, e alle competenze pratiche rappresentate dal protocollo ECDL (European Computer Driving Licence).²²

Nell'ambiente accademico italiano è sorto recentemente un meritorio corso di laurea in Informatica Umanistica presso l'università di Pisa,²³ che tuttavia consiste in un puro accostamento di discipline umanistiche ed informatiche; non c'è nemmeno un insegnamento dedicato all'informatica umanistica come tale. Allo stesso modo non vi è un "settore scientifico-disciplinare" legalmente riconosciuto, e non è prevista una reale valutazione delle ricerche in questo campo.

Su tutta questa materia sarebbe opportuno aprire una discussione seria, che possa avere dei riscontri nell'organizzazione universitaria, tenendo conto che l'ordinamento italiano, che è stato abbastanza liberalizzato nei corsi di studio (quello che si direbbe "a valle"), è tuttora rigido nella definizione degli insegnamenti che si impartiscono (quello che si direbbe "a monte"). Senza l'inclusione in questi insegnamenti di un gruppo specificamente dedicato all'informatica umanistica non si potranno ottenere risultati soddisfacenti sia per l'organizzazione della ricerca, sia per quella della didattica. I miei

sforzi in questa direzione – e sia pure per mia colpa e insufficienza – sono stati vani, e le prospettive non mi sembrano incoraggianti. Rimanderò dunque alle ragioni invocate in due saggi;

nota: Curriculum Fiormonte; Computerphilologie

noterò che recentemente si sono avuti in ambiente inglese due espliciti ed assai dettagliati riconoscimenti della disciplina, sia pure ciascuno con meriti e difetti;²⁴ e non insisterò oltre.

Merita piuttosto a questo punto menzionare i gruppi di discussione su e-mail ed i Congressi. I primi, ma soprattutto il benemerito Humanist²⁵ hanno svolto e tuttora svolgono un insostituibile compito di informazione e libera discussione fra i diretti interessati; i secondi, fra cui menzioneremo la serie annuale di Literary and Linguistic Computing²⁶ e quella pure ciclica di Fiormonte²⁷ sono invece ormai ridotti a stanche presentazioni di imprese più o meno interessanti nei diversi settori applicativi e svolgono più che altro un ruolo autoreferenziale.

In conclusione, in tutto questo agitarsi tecnologico, nel quale rilevo una notevole confusione, che sottolinea la mancanza di un quid disciplinare che aiuti a distinguere fra competenze umanistiche e competenze informatiche, fra aride applicazioni e attività computazionali, fra meriti di tipo scientifico e meriti di tipo applicativo, poche sono le attività propriamente definibili di ricerca. Ne proporremo due. La prima è quella che si esplica nel passaggio da quella che veniva chiamata linguistica computazionale, volta in sostanza a produrre i parser di analisi testuale di vario tipo, alla corpus linguistic, cioè alla linguistica sperimentale, così battezzata, a quanto mi risulta, da Geoffrey Sampson.²⁸

Essa utilizza la basi di dati testuali (corpora) per verificare in modo appunto sperimentale le teorie linguistiche basate sulla pura competenza grammaticale. In tal modo si sviluppa una interessante dialettica fra ricercatore e testo, ed una continua sperimentazione che rappresentano i migliori esiti dell'uso dell'informatica in ambito umanistico.

La seconda rimarchevole attività di ricerca è quella dei modelli archeologici, differenti dai GIS. È ancora troppo presto per emettere giudizi, ma la strada è sicuramente giusta. Modelli e sistemi²⁹ dovrebbero essere comunque la guida della ricerca in Informatica Umanistica, recuperando così in maniera virtuosa un ruolo della filosofia che i filosofi stessi trascurano nei loro rapporti con l'informatica. Notiamo soprattutto che non vengono prese sul serio la semiotica, la teoria dei sistemi, e la cibernetica, che certamente hanno dei grossi difetti, se assumono la pretesa di validità universale, ma hanno insegnato molte precisazioni concettuali essenziali per un corretto rapporto con i metodi informatici, in quanto distinti dalle pure applicazioni.

Si sta diffondendo invece una riflessione sull'Informatica Umanistica di stampo sociologico, centrata sull'interazione uomo-macchina, e sulle sue conseguenze per la mentalità che si forma nell'uso dei computer, soprattutto di internet, a causa della improvvisa disponibilità di grandi quantità d'informazione, peraltro di qualità non verificabile. È vero che tutto questo ha un interesse pragmatico per i produttori di software ed un interesse spicciolo per i lettori di giornali, ma non deve essere confuso con lo studio delle opportunità teoriche e metodologiche che l'informatica offre alle discipline umanistiche.

Allo stesso modo, dal punto di vista della ricerca umanistica, costituiscono quelli che chiamerei falsi obiettivi gli studi relativi ad archivi, biblioteche, editoria, didattica, nuovi media e multimedia, volti allo scopo di ottimizzare le prestazioni dei computer in questi settori. Essi sono di per sé altamente meritori, ma ancora una volta lo studioso umanista è per sua essenza interessato al contenuto e non alle forme che esso assume in ambiente informatico.

Si può parlare di prospettive? Che esse siano formidabili sotto l'aspetto della prassi e quello dell'apparenza non è chi non veda; per l'aspetto più intrinseco credo che ci sia solo da sperare che qualche giovane ricercatore prenda a cuore il lato teorico e metodologico dell'Informatica Umanistica e che possa trovare uno spazio nel quale poter operare.

1 Per una storia dell'Informatica Umanistica cf. <http://users.ox.ac.uk/~ctitext2/history/> -- Per le indicazioni bibliografiche (che qui omettiamo) relative ai singoli punti toccati rimandiamo a Giovanni ADAMO, *Bibliografia di Informatica Umanistica*, Roma, Bulzoni Editore, 1994 = *Informatica e Discipline Umanistiche* 5

2 cf. John Hutchins, *From First Conception to First Demonstration: the Nascent Years of Machine Translation, 1947 – 1954. A Chronology*, *Machine Translation* 12: 195-252, 1997.

3 http://it.wikipedia.org/wiki/Roberto_Busa.

4 <http://it.wikipedia.org/wiki/ENIAC>

5 http://it.wikipedia.org/wiki/Scheda_perforata#Perforatrici

6 <http://cognitivephilology.wordpress.com>

7 <http://en.wikipedia.org/wiki/SGML>

8 <http://www.w3.org/XML>

9 <http://www.twi.org>

10 <http://www.w4.org/html>

11 Vedi sotto per la sociologia.

-
- 12 <http://www.gis.com/whatisgis/index.html>
13 <http://ota.ahds.ac.uk>
14 <http://www.tlg.uci.edu>
15 LIZ 4.0, a cura di Pasquale Stoppelli ed Eugenio Picchi, Bologna, 2001.
16 <http://pespmc1.vub.ac.be/DEFAULT.html>
17 <http://www.humnet.unipi.it/cisiau>
18 http://www.unito.it/sezioni_federate/centri/C010.htm
19 <http://www.bronxlab.unimi.it> – soltanto servizi didattici.
20 <http://gandalf.aksis.uib.no/AcoHum>
21 Computing in Humanities Education, Univ. of Bergen, 1999, p. 13 sgg.
22 <http://www.ecdl.it>
23 infouma.di.unipi.it
24 Per un curriculum europeo di informatica umanistica, in D. Fiormente (ed.), Informatica umanistica. Dalla ricerca all'insegnamento, Roma : Bulzoni, 2003; Is Humanities Computing a Discipline?, «Jahrbuch für Computerphilologie» 4 (2002) 51-58
25 <http://www.princeton.edu/~mccarty/humanist> . Cf. anche Digital philology, Medievalist. Centernet.
26 <http://www.allc.org/content/conf/index.html>
27 http://www.cch.kcl.ac.uk/clip2006/content/practicalities/past_conferences-ita.html
28 Geoffrey Sampson, *Empirical Linguistics*, London and New York, Continuum, 2001
29 Cf. T. Orlandi, *I modelli fra informatica e tradizione*, in: Convegno internazionale sul tema *I modelli nella ricerca archeologica. Il ruolo dell'informatica*, Roma 23-24 novembre 2000, Roma, Accademia Nazionale dei Lincei, 2003, p. 165-176 (Contributi del Centro Linceo Interdisciplinare Beniamino Segre, 107)