

Introduzione

L'idea di riunire intorno allo stesso tavolo storici della scienza e studiosi di linguistica computazionale è stata proposta da Mauro Di Giandomenico in numerose occasioni: ad esempio nel dibattito che si sviluppò all'epoca del Congresso della Società Filosofica Italiana *L'uomo e la macchina. Trent'anni dopo*, tenutosi a Bari nel 1997.

Circa dieci anni dopo questo auspicio è divenuto realtà e, ancora a Bari, si è appunto tentato un dialogo fra specialisti di discipline relativamente lontane.

Come ricorda Giuliano Pancaldi nelle sue considerazioni conclusive, nella bibliografia internazionale di storia della scienza sono praticamente assenti le pubblicazioni che si occupano di questo incrocio disciplinare. In effetti, nonostante la storia della scienza riconosca da tempo fra i suoi utensili storiografici l'approccio quantitativo statistico inaugurato da Derek de Solla Price con la scientometria, le tecniche di *humanities computing* l'hanno finora interessata soprattutto sul fronte della catalogazione e della ricerca bibliografica, mentre rare sono state le applicazioni di *text mining* applicate a grandi *corpora* scientifici.

Dal canto suo, la linguistica computazionale è stata largamente applicata all'analisi di opere letterarie e poetiche ma raramente è stata vista come strumento efficace da parte degli storici della scienza.

Quest'ultimo fatto non è un caso, né è da ritenersi legato a caratteristiche tecniche particolari che impedirebbero di fatto l'uso degli algoritmi informatici nel caso dei testi scientifici. Questa condizione è invece da riportarsi presumibilmente ad una serie di pregiudizi degli studiosi delle varie discipline coinvolte:

- gli scienziati in generale ritengono che la scienza e la tecnologia risieda più negli oggetti che nei testi. In generale, come ha fatto notare Steve Fuller, a differenza degli umanisti, gli scienziati sono legati ai libri solo per l'apprendimento e per scopi educativi, ma non per fini di ricerca, per cui non portano un rispetto particolare per le forme letterarie di conoscenza conservate negli archivi, se si eccettua il loro essere a volte veicolo di "testimonianza virtuale" di fenomeni naturali o sperimentali. Questo per il fatto che la tecnologia non è un'impresa letteraria – e non perché inventori e ingegneri siano illetterati – piuttosto perché essi sono abituati a pensare in "oggetti" piuttosto che in "parole";
- gli storici della scienza di formazione filosofica, per converso, non nascondono le loro perplessità sulla utilità di approcci "meccanici" allo studio della storia delle idee, ritenendo a volte un'offesa alla propria capacità di comprensione l'affidamento ermeneutico ad una lettura fatta dal computer;
- infine, gli esperti di linguistica computazionale raramente compiono incursioni profonde in discipline nelle quali le loro tecniche possono essere applicate, preferendo restare saldamente ancorati allo sviluppo di procedimenti sempre più sofisticati all'interno del

proprio campo disciplinare.

Eppure, proprio nel campo della storia della scienza, l'applicazione di tecniche di *humanities computing* risulta doppiamente appropriata, perché in questo caso, il testo scientifico viene considerato in una duplice valenza:

1. come veicolo di idee, ed in quanto tale oggetto privilegiato da parte degli umanisti e degli storici. Secondo Willard McCarty, gli storici di solito hanno badato di più alle idee sulle cose che alle cose stesse. Di qui, la difficoltà che gli umanisti hanno a confrontarsi con ogni cultura intellettuale che si esprima altrimenti che con le parole. In altri termini, le *humanities* soffrono di *object-blindness*. E questo costituisce un ostacolo alla riassimilazione dello studio della cultura umanistica con quella della scienza, perché la seconda poggia appunto su molti apparati materiali. Ma questo ostacolo può essere in parte superato proprio dal *humanities computing* in quanto questa disciplina può considerare il testo scientifico...
2. ...come oggetto, appunto. Come *impressione*, nel senso di stampa di informazioni, e quindi oggetto storico-scientifico, cosa, nella quale ci sono idee al pari degli altri oggetti scientifici, come idee cioè sono contenute in una sfera armillare od in un pendolo foucaultiano. Ora, grazie alla linguistica computazionale abbiamo il mezzo per manipolare questi oggetti, armeggiare con essi grazie ai nostri computer al fine di ritrovare le tracce di quei paradigmi condivisi tacitamente, quella conoscenza di fondo, inespresa e secondo taluni inconoscibile e che invece proprio in questa manipolazione può tornare a farsi patente, come la manipolazione di de Solla Price condusse alla riattivazione dello scopo del Meccanismo di Antikythera.

Uno dei problemi più sentiti, quando si valuta l'impatto possibile delle tecniche della linguistica computazionale dei *corpora* (CCL) sulla storiografia scientifica è la scarsità di esempi concreti, di applicazioni pratiche, a fronte di una abbondante messe di proposte teoriche. Inoltre, alcuni programmi di ricerca operanti specificamente in questa direzione sembrano essersi arenati. È il caso – ad esempio – del progetto del CNRS francese *Colosciences (Corpus littéraires scientifiques)* che ambiva alla messa in linea di numerosi testi di biologi e naturalisti francesi del XIX secolo (in primo luogo Claude Bernard di cui si prevedeva la digitalizzazione di tutte le opere), con lo scopo di ricostruire una “storia delle idee biologiche” percorribile con criteri ipertestuali, tassonomici e semantici, mediati da tecniche linguistico-computazionali. Tuttavia, il sito non viene più aggiornato da cinque anni ed il progetto è rimasto incompiuto. Molto più completi, viceversa, gli Archivi Integrati di Risorse Galileiane realizzati dall'Istituto e Museo di Storia della Scienza di Firenze, che consentono anche una ricerca lemmatizzata all'interno dell'intero *corpus* galileiano.

Queste realizzazioni, come pure alcuni saggi comparsi nella collana di Storia della Scienza del Seminario barese, si possono catalogare nell'ambito dell'approccio detto *algoritmico* dell'informatica umanistica, vale a dire nell'ambito dei metodi basati sulla scansione del testo al fine della sua etichettatura grammaticale, della realizzazione dell'*index locorum*, della lemmatizzazione, della produzione di concordanze lemmatizzate, ecc. La debolezza di questo potente strumento informatico sta nel fatto che esso *non è pensato* per lo studio di *corpora* scientifici, e lo storico della scienza è costretto ad usare software progettato per altri scopi, adattandolo alle proprie esigenze disciplinari. Come il lettore di queste pagine vedrà, le cose vanno leggermente meglio con l'uso di metodi metalinguistici e, ancor più, rappresentazionali, che consentono un riarrangiamento del testo originale in grado di suggerire fertili ed originali percorsi interpretativi. Ne consegue la necessità che gli studiosi delle due discipline sconfinino, in un dialogo fecondo che è l'unico che potrebbe consentire ad entrambi l'invenzione, la produzione, la realizzazione di approcci efficaci per una linguistica computazionale storico-scientifica.

Nelle giornate baresi si sono ascoltate le posizioni, spesso ancora molto lontane fra di loro, di

storici della scienza, linguisti, linguisti informatici, teorici di *humanities computing*, informatici. Ovviamente, ne risulta una disomogeneità di base che però era nelle premesse ed è da considerarsi stimolante e, si spera, fruttuosa.

Bernardino Fantini, da storico della scienza, apre il convegno con il lancio di un programma di ricerca internazionale ed interdisciplinare volto allo studio del linguaggio delle emozioni nella storia della medicina e delle arti. È evidente che questo studio si gioverà in maniera essenziale delle tecniche di linguistica computazionale per studiare i mutamenti nel lessico emozionale nel tempo, a seconda dei campi disciplinari, in relazione alle concezioni fisiologiche (e filosofiche) di base. Lo studio di numerosi e vasti *corpora* di iatrofisici, musicisti, filosofi morali, permetterà l'applicazione delle tecniche metalinguistiche e rappresentazionali volte ad enucleare le metamorfosi dei dizionari – e delle semantiche sottese – nella storia dell'interpretazione delle passioni/emozioni. L'intervento di Willard McCarthy punta dritto al cuore del problema: perché l'uso dello strumento informatico, ormai adulto, non riesce ancora a produrre i risultati auspicati nel campo della critica letteraria e scientifica, risultati che sembravano vicinissimi da cogliere sin dalla fine degli anni settanta? L'argomentazione del filosofo britannico ruota intorno all'uso del testo e dell'informatica come *strumenti* in grado – nel loro reciproco rapporto – di produrre nuovi ed inaspettati livelli di complessità ed interpretazione. Quasi una risposta agli interrogativi sollevati da McCarty, il contributo di John Nerbonne che, dopo aver passato in rassegna appunto i tentativi falliti, le strade interrotte, le false partenze della linguistica computazionale applicata agli studi storico-filosofico-scientifici, individua nelle tecniche, da egli definite di “estrazione terminologica” e nel *bootstrapping* di tali tecniche che conduce alla costruzione di ontologie, la via per l'individuazione di porzioni di conoscenza scientifica “vera” estraibili in automatico dal testo. La questione da porre è se tale risultato non sia sovrapponibile a quello ottenuto dalla lettura fatta da un cervello umano degli stessi testi, o se – come noi crediamo – non sia in grado di rivelare anche conoscenze inespresse ed inaspettate. Un esempio di *work in progress* è quello riportato da Marco Veneziani sull'intero *corpus* delle opere di Cartesio che permetterà di percorrere l'intero apparato testuale del pensiero del filosofo con i classici strumenti della linguistica computazionale. Un saggio di quello che si potrà ottenere e una esposizione di quello che è già stato fatto con i classici metodi linguistici, appunto sul vocabolario cartesiano, è riportato nel meticoloso ed approfondito contributo di Franco Aurelio Meschini.

Con l'articolo di Alessandro Lenci entriamo nell'area di ciò che si fa oggi con l'obiettivo di applicare la linguistica computazionale allo studio di testi filosofici e scientifici. In ideale prosieguo con il saggio di Nerbonne, Lenci ci introduce – dopo un breve riflessione storica – nel vivo delle tecniche della produzione di ontologie mediante modelli distribuzionali. Vengono quindi mostrati in concreto i risultati ottenuti nella costruzione di “spazi semantici di parole” nei bruniani *Eroici Furori*. Con Lenci abbandoniamo dunque le questioni teoriche e fondazionali per addentrarci nella esplicazione dettagliata delle tecniche utilizzabili e nell'esposizione dei primi risultati concreti. La collocazione di questa concretezza e di questi risultati nel campo specifico della storia della scienza è l'obiettivo perseguito nelle tre relazioni finali dai ricercatori appartenenti al centro di ricerca interdipartimentale di Bari, il Seminario di storia della scienza. Lucia di Palo illustra i risultati conseguiti applicando il software di lemmatizzazione INTEX realizzato da Max Silberztein – che ha successivamente avuto numerose filiazioni e modifiche – alle *Recherches sur la vie et sur la mort* di Bichat. Scopo preciso della studiosa barese è la verifica di quanto originale sia riuscito ad essere il vocabolario bichatiano nel suo sforzo di riformare la fisiologia del XVIII secolo. E i risultati esposti mostrano elegantemente il parallelismo tra la controversa riuscita della riforma concettuale bichatiana e il “disagio definitorio” del suo lessico. Carla Petrocelli, da lungo tempo esperta nello studio linguistico computazionale del grande pisano, riporta le evidenze di titubanza ed incertezza di Galileo nel suo sforzo di costruire una nuova, efficace, retorica, lingua scientifica. Il contributo del sottoscritto, infine, dimostra l'applicazione delle metodologie rappresentazionali e le risultanze dei confronti fra spazi

semantici simili ma non identici nell'evidenziare sfumature epistemiche presenti nei testi di alcuni grandi fisiologi francesi del XIX secolo (Pierre Flourens, Claude Bernard, C. E. François-Franck), sfumature epistemiche che non sono sempre evidenti alla lettura non computazionale dei testi presi in esame.

Un'ultima notazione, per così dire di carattere storico, va aggiunta alle chiare ed efficaci conclusioni riassunte dall'intervento finale di Giuliano Pancaldi. Per quanto banale la seguente constatazione possa essere, sarebbe opportuno non sottovalutarla nello stimare quello che potranno essere gli scenari nei prossimi anni: l'aumento esponenziale delle tecniche e dei materiali disponibili. Quando padre Busa si dispose alla informatizzazione della *Summa dell'Aquinate*, più di cinquanta anni fa, lavorò con le schede perforate. La digitalizzazione e successiva lemmatizzazione dei tre tomi di carattere neurologico delle opere di J.-M. Charcot per una tesi di dottorato richiedeva tre anni ancora dieci anni fa. Oggi la messe di opere scientifiche in formato digitale disponibile sul web e su supporti ottici è enorme ed in continuo aumento giorno per giorno, mentre hardware sempre più potente e software sempre più sofisticati si offrono agli studiosi che vorranno utilizzarli. Esplorare – dal punto di vista di alcune ontologie semantiche – l'*Evolution créatrice* di Bergson in lettura parallela con *Le mouvement* di E. J. Marey costa, oggi, giusto due pomeriggi di lavoro. Non è detto che questo conduca necessariamente ad un mutato atteggiamento degli studiosi della storia delle idee scientifiche nei confronti della linguistica computazionale, tuttavia riteniamo sarebbe opportuno valutare almeno le possibilità offerte da questo vecchio ma nuovo strumento storiografico.

La realizzazione del workshop è stata resa possibile grazie al prezioso apporto di Istituzioni, Enti, studiosi e collaboratori. In particolare l'Università di Bari, nelle persone del Magnifico Rettore Corrado Petrocelli e del pro-Rettore Augusto Garuccio, l'Amplissimo preside della Facoltà di Lettere e Filosofia, Grazia Distaso, il Dipartimento di Scienze Filosofiche nella persona del Direttore Ferruccio De Natale e la Banca Popolare di Puglia e Basilicata.

Al workshop hanno preso parte come relatori e *chairmen* anche Mirella Conenna, Gerald Brey, Vincenzo Ambriola, Pasquale Guaragnella, Giovanni Semeraro. Hanno collaborato: Antonella Azzone, Betty Campanile, Enzo de Fazio, Rosanna Ficarella, Alessandro Volpone, Francesco Paolo de Ceglia, Caterina Tisci, Francesca Morgese, Sabrina Veneziani, Marusca Destino ed i miei allievi del corso B del IX ciclo SSIS di Bari.

Questo volume è stato dato alle stampe grazie al fondamentale contributo della Fondazione Cassa di Risparmio di Puglia.

Insostituibile l'apporto organizzativo di Filly Calabrese e Lucia di Palo senza il cui entusiasmo tutto questo sarebbe stato inattuabile.

L'idea del workshop e l'incitamento costante alla messa in opera sono merito di Mauro Di Giandomenico.