

Roberto Mazzola

Google Books e le scienze (post)umane



Laboratorio dell'ISPF, XII, 2015

DOI: 10.12862/ispf15L405
[Saggi - 5]

L'ipotesi di lavoro da cui prendo le mosse è che all'origine della crisi delle scienze umane di cui tanto si parla¹ vi sia il più complessivo mutamento epistemologico provocato dalle nuove tecnologie dell'informazione e della comunicazione (ICT). Quelle che avrebbero progressivamente eroso la «disposizione generale dell'*episteme*» che, secondo Foucault, a partire dal XIX secolo aveva consentito il costituirsi delle cosiddette scienze dell'uomo. Il filosofo francese individuava come caratteristica principale di queste ultime

non quell'oggetto privilegiato e singolarmente ingarbugliato che è l'uomo. Per la buona ragione che non è l'uomo che le costituisce e conferisce loro un campo specifico; è invece la disposizione generale dell'*episteme* che assegna ad esse un posto, le richiama e le instaura, consentendo loro in tal modo di costruire l'uomo come loro oggetto².

Secondo Foucault, l'uomo, in quanto soggetto/oggetto di autonomi saperi, è un costrutto storicamente determinato, un'invenzione relativamente recente della cultura europea, che può scomparire dalla scena così come è apparso.

Negli ultimi decenni del XX secolo la discontinuità epistemologica, di cui il filosofo francese poteva solo intravedere alcuni indizi, si è di fatto istituzionalizzata con l'affermazione del paradigma culturale che afferma, in ogni ambito della conoscenza, il valore fondante e unificante dei principi teorici delle tecnoscienze computazionali e informazionali.

L'estensione agli studi umanistici di questo nuovo approccio alla realtà e alla conoscenza ha suscitato le resistenze di quanti hanno difficoltà ad accettare l'idea di ridurre un libro, un dipinto, un brano musicale ecc., a mero flusso di informazioni codificate. Ciò nonostante, dagli anni ottanta del secolo scorso, il ridimensionamento istituzionale della "provincia del sapere" occupata dalle discipline umanistiche, le ha relegate ai margini culturali dell'«impero cibernetico»³, in cui Jerome Lanier, pioniere della realtà virtuale, vede profilarsi il rischio dell'affermarsi di un vero e proprio «totalitarismo cibernetico»⁴. Considerata in questo contesto, la riflessione sullo stato attuale e soprattutto sul futuro delle scienze umane, al di là di ogni altra considerazione, non può limitarsi a prendere atto dell'incidenza sempre più massiccia della mediazione tecnologica nelle discipline umanistiche, nel momento in cui ad essere messo in discussione è lo

¹ M. C. Nussbaum, *Non per profitto. Perché le democrazie hanno bisogno della cultura umanistica*, Bologna, Il Mulino, 2013. All'appassionata difesa di Martha C. Nussbaum dell'attualità dei valori e dell'istruzione umanistica classica per l'educazione alla cittadinanza globale democratica delle future generazioni replica R. Braidotti, *Il postumano. La vita oltre la specie, oltre la morte*, Roma, DeriveApprodi, 2014, che (nel capitolo 4) rivendica l'ottimo stato di salute delle scienze umane che abbandonando la visione liberale umanista affrontano le problematiche dell'attuale condizione postumana.

² M. Foucault, *Le parole e le cose. Un'archeologia delle scienze umane*, Milano, Rizzoli, 1978, p. 390.

³ Così Céline Lafontaine ha definito la costruzione sociale nata dalla rivoluzione digitale : cfr. C. Lafontaine, *L'Empire cybernétique. Des machines à penser à la pensée machine*, Québec, Éditions du Seuil, 2008.

⁴ J. Lanier, *One Half of a Manifesto*, 11 ottobre 2000, <https://edge.org/conversation/jaron_lanier-one-half-a-manifesto>; Id., *Tu non sei un gadget, Perché dobbiamo impedire che la cultura digitale si impadronisca delle nostre vite*, Milano, Mondadori, 2010.

stesso fondamento epistemologico che ha guidato la ricerca scientifica in età moderna.

Nel 2008, dando voce allo stile di pensiero ipertecnologico *mainstream* nella Silicon Valley, Chris Anderson, caporedattore dal 2001 al 2012 di «Wired» – rivista nota come la bibbia dei *nerds* – ha provocatoriamente proclamato «la fine della teoria»⁵. In estrema sintesi, secondo il giornalista e saggista statunitense, nell'era di *Google* e dei *Big Data*⁶, ovvero dei miliardi di informazioni inserite in un database e sottoposte a procedure algoritmiche, il metodo sperimentale praticato dalla scienza moderna basato sulla triade ipotesi-modello-verifica, è ormai obsoleto, e va sostituito dal binomio dati/algoritmo che è in grado di scoprire le correlazioni nascoste della realtà senza dover ricorrere a esperimenti di laboratorio, elaborazione di modelli e teorie⁷.

⁵ C. Anderson, *The End of Theory: The Data Deluge makes the Scientific Method Obsolete*, in «Wired», 23 June 2008, <<http://www.wired.com/2008/06/pb-theory/>>. Il 30 giugno del 2008 The Reality Club ha ospitato le considerazioni “a caldo” di George Dyson, Kevin Kelly, Stewart Brand, W. Daniel Hills, Sean Carroll, Jaron Lanier, Joseph Traub, John Horgan, Bruce Sterling, Douglas Ruschoff, Oliver Morton, Daniel Everett, Gloria Origgi, Leo Smolin e Joel Garreau: cfr <<http://edge.org/discourse/discourse.html>>.

⁶ Il termine “Big Data” è diventato popolare dopo il caso Snowden oggetto dell'appassionante libro di G. Greenwald, *Sotto controllo. Edward Snowden e la sorveglianza di massa*, Milano, Rizzoli, 2014. La regola delle tre V (volume, velocità e varietà) sintetizza la strategia di marketing dei principali produttori di servizi in Big Data (IBM, Teradata e EMC) offerti ai soggetti pubblici e privati che intendono sfruttare l'aumento esponenziale delle informazioni digitalizzate e presenti in Rete. Per un'approfondita analisi critica vedi R. Kitchin, *The Data Revolution. Big Data, Open Data Infrastructures and their Consequences*, London, SAGE Publications, 2014. Nell'accezione più ampia Big Data «designa quelle cose se si possono fare solo su larga scala, per estrapolare nuove indicazioni o creare nuove forme di valore, con modalità che vengono a modificare i mercati, le organizzazioni, le relazioni tra cittadini e governi, e altro ancora» (V. Mayer-Schönberger - K. Cuchier, *Big Data. Una rivoluzione che trasformerà il nostro modo di vivere – e già minaccia la nostra libertà*, Milano, Garzanti, 2013, p. 16). Non è la prima volta nella storia che si pone il problema del sovraccarico di informazioni: basti pensare al supporto offerto agli studiosi dagli eruditi bibliotecari dell'antica biblioteca di Alessandria o all'impegno profuso dai letterati europei del Rinascimento nel redigere indici, bibliografie, enciclopedie considerati indispensabili strumenti di base per la ricerca e il recupero delle informazioni utili non solo all'originale rielaborazione del ritrovato patrimonio culturale dell'antichità classica ma anche per organizzare le nuove conoscenze naturalistiche e antropologiche accumulate nell'età delle scoperte geografiche.

⁷ La classica obiezione che correlazione non implica causazione non viene presa in considerazione per il semplice motivo che, per Anderson, la relazione fra le variabili è una scatola nera che è inutile tentare di aprire. Si tratta di un approccio ampiamente utilizzato nella recente letteratura scientifica che raccoglie enormi quantità di dati e evidenze empiriche per mostrare come il fattore x è collegato al fattore y senza spiegarne il motivo. Ma l'aspetto più inquietante è che la verifica della significatività statistica dei risultati, per la complessità tecnologica che li ha generati, è lunga, laboriosa e costosa. Le lacune nel sistema delle pubblicazioni accademiche, meccanismo fondamentale del processo di conoscenza scientifica, sono state messe in luce da John Ioannidis in uno studio del 2005 dal titolo significativo: *Perché la gran parte delle evidenze scientifiche pubblicate sono false* (J. Ioannidis, *Why Most Published Research Findings Are False*, <<http://robotics.cs.tamu.edu/RSS2015NegativeResults/pmed.0020124.pdf>>). L'indagine dell'epidemiologo di Stanford, basata su sull'analisi statistica inferenziale della letteratura biomedica, mostra come l'enorme aumento dei ricercatori e delle pubblicazioni renda estremamente

Com'era prevedibile, le tesi *tranchant* di Anderson non potavano non suscitare la reazione di numerosi intellettuali, scienziati e addetti lavori, che hanno visto nella «fine della teoria» una riproposizione del vecchio sogno empirista-positivista, già demolito dal dibattito epistemologico novecentesco, di sottomettere l'impresa scientifica ad una presunta auto-evidenza dei dati, nel nome di una scienza neutrale svincolata da modelli esplicativi, ipotesi teoriche e contesto storico⁸.

Se da un lato Anderson sottolinea i successi dei *Big Data* nei programmi di ricerca delle scienze “dure” che adottano il modello organizzativo e finanziario della *big science*, dall'altro suona le campane a morto per gli scienziati sociali⁹. Forse nella scia dei successi, poi rivelatisi effimeri, di *Google Flu Trends*¹⁰, Anderson proclama:

difficoltosa la replica degli esperimenti e la falsificazione dei risultati la cui validità viene spesso accolta acriticamente non solo dall'opinione pubblica ma anche dagli esperti. Le questioni sollevate da Ioannidis rimandano ad un modello di scienza che si auto-coregge evidentemente estraneo alla cornice mentale di Anderson che ritiene che gli scienziati possano finalmente smettere di formulare ipotesi basate sulle osservazioni, costruire modelli, elaborare teorie da sottoporre al vaglio dagli esperimenti per affidarsi fiduciosi ai risultati dell'analisi statistica automatica dei Big Data che sono intrinsecamente veritieri e comprensibili da chiunque abbia dimestichezza con la visualizzazione grafica computerizzata.

⁸ In realtà l'empirismo radicale *data driven* di Anderson differisce non poco da quello «classico». Infatti, all'obiezione che ogni dato è già sempre il risultato di un'interpretazione lo scienziato empirista avrebbe ironicamente risposto: «È vero, non esistono fatti ma solo interpretazioni. E questo è un fatto!». L'empirismo, pur tracciando una netta linea di demarcazione tra il momento teorico e l'osservazione fattuale, non nega il ruolo dell'agente umano nel processo conoscitivo; per Anderson invece il momento soggettivo dell'elaborazione teorica è semplicemente un'inutile finzione di cui oggi si può fare a meno. Anche se Anderson annuncia la “fine della teoria” il suo discorso è epistemologicamente fondato sulla teoria dell'informazione che dal punto di vista matematico e ingegneristico considera irrilevante il senso e il significato del messaggio trasmesso. Per “leggere” le informazioni contenute nei Big Data le conoscenze informatiche prevalgono sulle competenze specialistiche ad esempio in biologia o in chimica. Grazie ai principi matematici della teoria dell'informazione, all'elettronica e alla potenza di calcolo ogni scienza diventa un'info-scienza *data driven*. Una ricognizione critica del dibattito sulle implicazioni epistemologiche della “fine della teoria” di Anderson in R. Kitchin, *The Data Revolution*, cit, cap. VIII («The Reframing of Science, Social Science and Humanitis Reserch»).

⁹ A differenza di quanto pensa Anderson, un uso intelligente del mix tra nuove tecnologie e *expertise* umano ci viene delle scienze sociali dove sempre più spesso alle tradizionali tecniche statistiche si affiancano software dedicati di *media monitoring* e *sentiment analysis* dell'evoluzione dei fenomeni sociali attraverso i social media. Interessanti esempi di queste applicazioni in Italia in A. Ceron - L. Curini - S. M. Iacus, *Social Media e Sentiment Analysis. L'evoluzione dei fenomeni sociali attraverso la Rete*, Milano, Springer, 2013; A. Lorenzet, *Il lato controverso della scienza. Nanotecnologie, biotecnologie e grandi opere nella sfera pubblica*, Bologna, Il Mulino, 2013.

¹⁰ Software di analisi dati nato dalla collaborazione dei ricercatori di Google con il centro di controllo e prevenzione delle malattie degli Stati Uniti. L'obiettivo era di utilizzare le ricerche degli utenti di Google per monitorare e il diffondersi delle epidemie influenzali. Nel 2007-2008 i risultati furono sorprendenti: il confronto tra i risultati di Google e quelli dell'organismo pubblico mostravano una concordanza del 97 per cento. L'iniziale entusiasmo degli ultraliberisti per la capacità di Google di fare previsioni, a riprova della presunta superiorità del privato sul pubblico, si è poi, per così dire, “raffreddato” dopo che nei successivi tre anni Google Flu Trends ha sovrastimato i casi di influenza fino a prevederne il doppio di quelli

Questo è un mondo in cui enormi quantità di dati e di matematica applicata sostituiscono ogni altro strumento che si possa immaginare. Abbandonate ogni teoria sui comportamenti umani dalla linguistica alla sociologia. Dimenticate tassonomia, ontologia e psicologia. Chi sa perché la gente fa quello che fa? Lo fa e basta! Siamo in grado di tracciarlo e misurarlo con una fedeltà senza precedenti. Con abbastanza dati, i numeri parlano da soli¹¹.

Lasciando da parte l'abusato espediente retorico dei numeri che parlano da soli e l'allusione ai deludenti risultati del web semantico auspicato da Tim Berners-Lee, il modello di scienziato sociale proposto da Anderson è quello del *data scientist*, che nello studio delle dinamiche di processi sociali, politici e culturali utilizza gli stessi strumenti di *data mining* e monitoraggio dei *social media* applicati al marketing delle imprese commerciali.

Anche se gli umanisti non sono neanche presi in considerazione da Anderson, vale la pena sottolineare che, nella prospettiva da lui indicata, nell'era di *Google* e dei *Big Data*, la disponibilità di risorse digitali e di analisi computazionali su scala finora impensabile renderebbe anacronistica non solo la figura dell'umanista tradizionale, ma anche di quello "digitale". L'introduzione della ricerca *data driven* nell'ambito delle scienze umane rappresenta una sfida di natura metodologica ed epistemologica, prima che tecnologica, che non può essere elusa da quanti considerano la tecnologia e la cultura digitale una risorsa e non un pericolo per le scienze umane. In questa prospettiva, per gli umanisti del XXI secolo diventa sempre più urgente elaborare nuove strategie culturali per un uso critico e alternativo delle ICT.

Come è noto, negli ultimi anni l'umanistica digitale, nella sua più ampia accezione inglese di *Digital Humanities*, ha conosciuto una rapida espansione, e anche se l'accostamento dei due termini può apparire un ossimoro, vista l'intrinseca natura analogica degli oggetti di studio e degli strumenti d'indagine delle discipline umanistiche, per ironia della sorte è proprio in quest'ambito che l'informatica ha trovato le sue prime applicazioni al di fuori dell'uso a scopi militari e scientifici. Come è stato correttamente osservato, nel rapporto tra informatica e scienze umane possiamo distinguere due approcci: «la computazione per le scienze umane (impostazione strumentale) e la computazione nelle scienze umane (impostazione metodologica)»¹². Si tratta di approcci anche cronologicamente distinti.

Il primo ha caratterizzato fin dalle origini l'informatica umanistica. La storia è nota. Nel 1949 il padre gesuita Roberto Busa riesce a convincere i dirigenti dell'IBM a fornire i finanziamenti, ma soprattutto i computer necessari per la

effettivamente registrati. L'indagine della rete era sì efficace nel rilevare malattie di tipo parainfluenzale ma era di scarso utilità per decidere il momento in cui avviare i programmi di vaccinazione. Ciò ovviamente non vuol dire che i Big Data siano inaffidabili, tutt'altro: in generale possiamo dire che i Big Data in genere funzionano abbastanza bene, soprattutto quando sono applicati al dominio ristretto di problemi ben definiti.

¹¹ C. Anderson, *The End of Theory*, cit.

¹² F. Bennato, *Il computer come macroscopio. Big data e approccio computazionale per comprendere i cambiamenti sociali e culturali*, Milano, Franco Angeli, 2015, p. 89.

lemmatizzazione delle opere di Tommaso d'Aquino. Il lavoro dura trent'anni e si adegua ai cambiamenti dei supporti di registrazione: passando dalle schede perforate ai nastri magnetici, ai floppy disk ecc. La passione e la competente tenacia di Busa vengono premiate con la realizzazione, nel 1989, dell'*Index Thomisticus* interattivo su CD-ROM, messo poi in rete nel 2005. Inutile dire che padre Busa, il quale non rinnegava la tradizione, appronta un'edizione a stampa dell'opera in 56 volumi. Come sempre accade ai pionieri, il religioso gesuita aveva aperto una strada che in molti avrebbero seguito, a partire dal progetto Gutenberg dei primi anni Settanta.

Negli ultimi due decenni del secolo scorso, l'introduzione e diffusione dei personal computer prima e del *Web* poi ha consentito agli studiosi, ai bibliotecari e ai bibliofili appassionati di tecnologia, di far circolare nel mondo accademico e non versioni elettroniche dei classici antichi e moderni non coperti da diritti d'autore; dagli anni novanta alcuni editori avviarono perciò la distribuzione digitale dei titoli più venduti del loro catalogo. Al principio del nuovo millennio, l'informatica umanistica si apriva a nuove forme di conciliazione delle metodologie tradizionali con le tecnologie digitali, tali da superare i modelli puramente lessicografici della linguistica computazionale, sperimentando nuovi modi di pensare e lavorare nella realtà ipertestuale e multidimensionale della rete.

Tuttavia, a fronte della diffusione in ambito accademico dell'uso di risorse informatiche negli studi letterari, sociali e culturali¹³, si è diffusa tra molti ricercatori una sorta di rassegnazione alla colonizzazione digitale della cultura umanistica, nella convinzione che l'uso del suffisso -info sia ormai indispensabile per ottenere finanziamenti, da utilizzare, anche se solo in parte, per organizzare in piena autonomia tempi e modi dei propri studi. Un'illusione di sopravvivenza che rende estremamente attuale il suggestivo parallelo, fatto da Lyotard nel *lontano* 1987, tra la decisione degli ebrei di contribuire alla costruzione del muro del ghetto di Varsavia nel quale avrebbero trovato la morte, e gli scrittori e i pensatori che

se resistono all'uso oggi predominante del tempo, non sono solo predestinati a scomparire, devono ancora contribuire alla fabbricazione di un "cordone sanitario" che li isoli. Al riparo di questa protezione, si presume che la loro distruzione possa essere differita. Ma essi "acquistano" questa proroga di sopravvivenza breve e vana, modificando la loro maniera di pensare e di scrivere in modo che le opere divengano più o meno comunicabili, scambiabili, in una parola commercializzabili¹⁴.

Comunque sia, ad un primo provvisorio bilancio degli effetti degli sviluppi della rivoluzione digitale, ritengo giusta e condivisibile l'opinione di Domenico Fiormonte, per il quale «le *digital humanities* in versione globalizzata, fino ad og-

¹³ Database, repertori, librerie e riviste digitali, digitalizzazione ed edizioni critiche di testi realizzate seguendo le linee guida della Text Encoding Initiative.

¹⁴ J.-F. Lyotard, *L'umano. Divagazioni sul tempo*, Milano, Lanfranchi Editore, 2001, p. 107.

gi, non hanno né rafforzato le scienze umanistiche né riequilibrato i rapporti di forza fra queste e le scienze “dure”¹⁵.

Vorrei a questo punto provare ad analizzare l'attuale situazione di stallo alla luce dell'ingresso di un colosso come *Google* nel campo dei beni culturali.

C'era una volta il libro stampato in tipografia, poi arrivò *Google Books* e nulla fu più come prima. Vero è che *Amazon*, quando era ancora un semplice rivenditore di libri in Rete, offriva ai potenziali clienti la funzione *Look Inside* per visualizzare l'indice e piccole parti di testo, ma il progetto avviato nel 2002 da *Google* aveva qualcosa di visionario che non poteva lasciare indifferenti, anche se c'era chi considerava l'iniziativa velleitaria se non impraticabile. Uno scetticismo destinato a sciogliersi come neve al sole via via che decine di biblioteche seguirono l'esempio delle prime cinque associate al progetto (le università di Harvard, di Stanford, del Michigan, di Oxford e la Public Library di New York). Per chi ritiene che il sapere sia un bene comune c'era di che rallegrarsi, e la messa in rete della conoscenza tramandata dai libri sembrava finalmente dare ragione a Francis Bacon che quattro secoli orsono perentoriamente affermava che «uno solo è lo stato delle scienze e questo fu e sarà sempre democratico»¹⁶.

Come tutti sanno, i libri digitalizzati da *Google* sono consultabili e scaricabili per intero gratuitamente, quando non siano coperti da *copyright*. A oggi, anche se le cifre oscillano, pare che ne siano stati già stati archiviati circa 30 milioni dei circa 130 editi dall'invenzione della stampa¹⁷. L'impatto su analoghe iniziative di editoria e biblioteche digitali è stato deflagrante, e oggi *Google* sembra non avere più competitori né privati né pubblici. Nel 2008 *Microsoft* ha chiuso *Live Search Books* lasciando però accessibili circa un milione di libri nel sito *Internet Archive*. Per tentare di contenere l'espansione del colosso americano, nel 2005 l'Europa rispondeva con il lancio di un progetto comunitario: l'*European Digital Library*, mentre la Biblioteca Nazionale di Francia dava vita a *Gallica*.

La lodevole intenzione dell'Europa di creare una biblioteca digitale comune senza scopo di lucro che mettesse a disposizione di chiunque a titolo gratuito la memoria collettiva del vecchio continente ha avuto non poche difficoltà a concretizzarsi ma soprattutto a svilupparsi vista la determinazione con la quale *Google* ha portato avanti il suo progetto di biblioteca universale. Di fronte all'impari disponibilità di risorse economiche e in mancanza di una efficace strategia comune, l'Europa ha infine capitolato. Negli ultimi anni, accordi di partenariato sono stati siglati da importanti biblioteche europee. In Italia, ad esempio, il 9 marzo del 2010 è stato raggiunto un accordo tra il MIBAC e *Google* che coinvolge le biblioteche nazionali di Roma, Firenze e Napoli oltre all'Istituto Italiano per il Catalogo Unico (ICCU). Ma il simbolo della resa eu-

¹⁵ D. Fiormonte, *Per una critica culturale delle Digital humanities*, in «Quaderni DigiLab», II, 2012, p. 236.

¹⁶ F. Bacon, *La grande instaurazione*, in Id., *Scritti filosofici*, a cura di P. Rossi, Torino, Utet, 1986, p. 523.

¹⁷ Per la precisione, fino al 5 agosto 2010 *Google* ne ha censiti 129.864880: vedi, sul blog di *Google* del 5 agosto 2010, *You can count the number of books on the world on 25,972,976 hands*.

ropea è sicuramente l'accordo siglato nel 2013 tra il Ministero della Cultura francese e l'azienda americana, che prevede che i documenti saranno d'ora in poi consultabili solo nella sede della BNF e non più da remoto, e soprattutto che i testi digitalizzati solo dopo dieci anni saranno messi *on-line* su *Gallica*. Anche se in rete continuano ad essere disponibili un centinaio di biblioteche digitali, la prima ad essere consultata sarà sempre, inesorabilmente, *Google Books*, che di fatto ha cannibalizzato tutti i concorrenti.

L'aspirazione di *Google* al controllo monopolistico dei contenuti digitali, dei servizi di ricerca e dei diritti di accesso sfrutta abilmente l'irresistibile fascino di avere a portata di *click* il sapere e il piacere estetico racchiuso in tutti i libri del mondo. In particolare, per gli studiosi abituati a leggere libri pubblicati fino al XIX secolo, i testi di *Google Books* rappresentano una sorta di "anastatiche elettrificate", copie digitali di libri da leggere dopo averli stampati, e più di rado direttamente sullo schermo del computer. Inoltre, poiché *Google* tratta l'immagine del testo con avanzati *software* per il riconoscimento dei caratteri a stampa (OCR *omni-font*) di oltre 35 lingue, è possibile individuare le parole chiave o le porzioni di testo che interessano; si tratta di un uso strumentale del supporto digitale che all'apparenza non sembra differire molto dal tradizionale approccio qualitativo della lettura attenta, dello studio paziente e scrupoloso di un insieme limitato di opere.

L'aver reso disponibile uno strumento che rende la vita più facile a chi è abituato a frequentare libri e biblioteche suscita uno spontaneo sentimento di ammirazione e gratitudine nei confronti dello spirito filantropico di *Google*, nonostante le obiezioni sollevate da alcuni autori come, ad esempio, Nicholas Carr, che ha sottolineato come per *Google*

il vero valore di un libro non [sia] quello di un'opera letteraria compiuta in se stessa, quanto piuttosto quello di un ulteriore ammasso di dati in cui scavare. La grande biblioteca che *Google* si sta affrettando a creare non deve essere confusa con le biblioteche che abbiamo conosciuto fino a oggi. Non è una biblioteca di libri, è una biblioteca di frammenti¹⁸.

Vale a dire una biblioteca di *snippets*, lacerti di conoscenza messi a disposizione degli appassionati di citazioni di seconda mano. Il valore di un libro invece, sottolinea Carr, risiede nella sua compiutezza, che per essere compresa e apprezzata necessita di una lettura diretta non finalizzata a raccogliere solo informazioni¹⁹.

¹⁸ N. Carr, *Internet ci rende stupidi? Come la Rete sta cambiando il nostro cervello*, Milano, Raffaello Cortina, 2011, p. 199. Di tutt'altra opinione è H. Rheingold, *Perché la rete ci rende intelligenti*, Milano, Raffaello Cortina, 2012. Il libro vuole essere una vera e propria guida pratica all'uso intelligente e consapevole del Web. In Italia lo scrittore ed editore Roberto Calasso (*L'impronta dell'editore*, Milano, Adelphi, 2013) nel fare l'apologia dell'editoria tradizionale ha condannato senza appello l'idea stessa di digitalizzare i libri.

¹⁹ A questa obiezione l'azienda risponde scaricando la responsabilità su autori ed editori. Se chiediamo «Perché non posso consultare tutto il libro?», *Google Books* ci informa che «Molti dei titoli presenti su *Google Libri* appartengono agli autori ed editori che partecipano al Pro-

In ogni caso, di fronte al successo dell'iniziativa non ci sono obiezioni che tengano perché, come dichiarato da un manager dell'azienda:

è ovvio che è una buona idea quella di rendere disponibile per il maggior numero possibile di persone la maggiore quantità di informazione possibile. È ovvio che aiutare la gente a trovare queste informazioni è un bene. Ed è ovvio anche che un autore trae beneficio solo se la gente trova che il suo libro contiene informazioni utili. Non ci sono sfumature di grigio su questo punto. La verità, in fin dei conti, è una funzione binaria²⁰.

«Ovvio», si potrebbe obiettare, se si vede il mondo attraverso le lenti degli 0 e degli 1 del codice binario; ma non per quanti hanno visto in *Google Books* non tanto un'opportunità quanto piuttosto una minaccia per il futuro dei libri e delle biblioteche. In particolare, negli Stati Uniti si è sviluppato un acceso dibattito intorno alle possibili conseguenze negative per gli editori e ai rischi di privatizzazione della gestione delle biblioteche. «Cosa succederà se *Google* preferirà la redditività all'accesso?»²¹. Del resto, è ben noto che una delle strategie del marketing digitale è di offrire inizialmente gratis servizi che poi diventano a pagamento.

Nel 2005 il sindacato degli scrittori e l'associazione degli editori statunitensi hanno citato in giudizio *Google*. Gran parte del motivo del contendere ruotava intorno agli interessi economici legati alle opere digitalizzate coperte da diritto d'autore (che negli Usa sono quelle pubblicate dopo il 1923) e delle opere fuori commercio. Come tutti i contenziosi a sfondo economico, alla fine le parti hanno raggiunto nel 2008 un primo accordo extragiudiziale con il cosiddetto *Google Book Settlement*²². L'accordo però non ha posto fine alla battaglia legale

gramma Partner. Di conseguenza, sono i partner a decidere quale percentuale dei propri libri rendere disponibile, da poche pagine di esempio al testo completo. Per quanto riguarda le opere che entrano a far parte di Google Libri tramite il Progetto Biblioteche, la parte disponibile dipende dallo stato del copyright. Rispettiamo le leggi sul copyright e lo straordinario sforzo creativo degli autori. Se il libro è di dominio pubblico e quindi non più protetto da copyright, è possibile sfogliarlo tutto e persino scaricarlo per poterlo leggere senza visitare il sito. Se, invece, il libro è protetto da copyright e l'editore o l'autore non hanno aderito al Programma Partner, ne vengono messe a disposizione esclusivamente le informazioni di base in una struttura simile a un catalogo a schede. In alcuni casi viene visualizzato anche qualche frammento, ovvero delle frasi che contengono il termine di ricerca nel contesto. L'obiettivo di Google Libri è aiutare gli utenti a trovare i libri e sapere dove acquistarli o prenderli in prestito, non quello di permettere loro di leggerli online dall'inizio alla fine. È come andare in libreria e sfogliare i libri, ma nello stile di Google».

²⁰ Cit. in E. Morozov, *Internet non salverà il mondo*, Milano, Mondadori, 2014, p. 162.

²¹ È la domanda che si è posto Robert Darnton, direttore del sistema bibliotecario di Harvard, dando voce ai bibliotecari che denunciavano i pericoli per la libera circolazione della cultura una volta che i libri spariranno dalle biblioteche sostituiti dalle loro copie digitali distribuite da un'azienda privata (R. Darnton, *Google and the Future of Books*, in «New York Review of Books», 12 febbraio 2009).

²² A. De Robbio, *La gestione dei diritti nella digitalizzazione di massa. Un'analisi alla luce del caso Google Book Search*, in «Bibliotime», XII, 2009, 2 (<<http://eprints.rclis.org/13506/1/GoogleLibri7marzo2009.pdf>>).

della *Authors Guild*, che si è rivolta alla giustizia ordinaria che, con sentenza emessa nell'ottobre del 2015 dalla seconda corte d'appello di New York, ha stabilito che *Google* non viola le leggi sul *copyright*. Una sentenza salutata da *Google* come una vittoria non tanto per l'azienda quanto per la libera circolazione del sapere. Non si può fare a meno di ricordare la ben più triste sorte toccata ad Aaron Swartz, arrestato nel luglio del 2011 per avere scaricato l'intero l'archivio di articoli accademici contenuto nel database JSTOR del MIT di Cambridge (Ma), accessibile a pagamento. Anche per Aaron, rendere la conoscenza disponibile gratuitamente a tutti era un obbligo morale, ma nel suo caso per aver commesso questo "crimine" rischiava una condanna fino a 35 anni di carcere. Nel gennaio del 2013, all'età di 26 anni, Aaron Swartz si è suicidato²³.

Molte delle critiche alla privatizzazione della conoscenza sono giuste e condivisibili, come anche comprensibile è la difesa degli interessi economici degli editori e degli autori; ma non credo che l'ingente investimento economico di *Google* sia esclusivamente finalizzato ad un ritorno in termini economici. Visti i numerosi progetti portati avanti nei laboratori di *Google X*, ritenere che il colosso americano sia solo un motore di ricerca che fa soldi con la pubblicità è quanto meno ingenuo; anche perché, come hanno scritto Erich Schmidt e Jonathan Rosemberg, probabilmente da qualche parte del mondo c'è già all'opera un gruppo di «creativi smart» impiegati a creare una società che renderà *Google* irrilevante. Un destino che, per quanto possa oggi apparire assurdo, è però inevitabile «dato che nessun business vince per sempre»²⁴.

Come tutte le aziende quotate in borsa, *Google* punta a incrementare il valore delle sue azioni, ma si vanta di seguire la regola 70/20/10 nell'allocatione delle risorse: cioè 70 per cento degli investimenti impiegati per sviluppare il motore di ricerca e la raccolta pubblicitaria, 20 nei prodotti emergenti e 10 investito in idee completamente nuove²⁵.

Forse nel 2002, nelle intenzioni di Larry Page, *Google Books* più che essere una macchina per fare soldi era solo un progetto avventuroso esposto a un grande rischio di fallimento. La vulgata tramandata dalla *Google Books History* narra di Larry Page e Marissa Mayer intenti il primo a fotografare, la seconda a voltare pagina di un libro acquistato su una bancarella, cronometrando il tempo necessario per la riproduzione²⁶. Al di là della retorica aziendale che vuole farci credere che *Google Books* sia una delle tante attività di responsabilità sociale (*corporate social responsibility*), in genere finanziate dalle multinazionali coi soldi accantonati con l'elusione fiscale, con il passare degli anni è apparso chiaro che in realtà la digitalizzazione dei libri era solo il primo abbozzo di un progetto culturale più ambizioso che Siva Vaidhayanathan ha definito la googlizzazione

²³ Sull'emblematica vicenda del giovane programmatore e attivista vedi *Aaron Swartz (1986-2013). Una vita per la cultura libera e la giustizia sociale*, all'URL <<http://aaronswartztributo.tumblr.com/>>.

²⁴ E. Schmidt - J. Rosemberg, *Come funziona Google*, Milano, Rizzoli, 2014, p. 334.

²⁵ Ivi, pp. 283-287. La prefazione di Larry Page, cofondatore di Google, dà l'*imprimatur* dell'ufficialità alla filosofia aziendale dagli autori.

²⁶ Disponibile in rete all'URL <<http://www.google.com/googlebooks/about/history.html>>.

della conoscenza²⁷. Un ulteriore passo avanti in questa direzione è stata la nascita, nel 2011, del *Google Cultural Institute*, che offre visite online di siti archeologici, musei, *World Wonders*, oltre che foto e materiale documentario provenienti da istituzioni o singoli utenti disposti a collaborare con *Google*. Il *Cultural Institute* ha ovviamente dedicato una particolare attenzione al Bel Paese, oggetto dell'accurata indagine di Elisa Bonacini²⁸. Come per la digitalizzazione dei libri, anche in questo caso l'obiettivo dichiarato è quello di rendere accessibili l'arte e la cultura, e al contempo dimostrare di essere l'unico soggetto in grado di gestire la cultura digitale nell'era della Rete; non è forse casuale che nel 2013, a ridosso dell'inaugurazione del *Cultural Institute*, l'azienda abbia prodotto il documentario *Google and the World Brain*²⁹, che già dal titolo evoca la matrice culturale e l'immaginario scientifico che animano la politica culturale del colosso californiano. *World Brain* è infatti il titolo della raccolta di conferenze e interventi tenuti tra il 1936 il 1938 da H.G. Wells, in cui lo scrittore britannico auspicava la creazione di una *Permanent World Encyclopedia* in grado di gestire il crescente aumento delle conoscenze. Le idee di Wells furono presentate al Congresso mondiale della documentazione universale di Parigi del 1937, al quale lo scrittore partecipò come delegato del Regno Unito, e rielaborate nell'autunno dello stesso anno in una serie di pubbliche letture tenute negli Stati Uniti, in cui Wells precisava le caratteristiche del nuovo enciclopedismo da lui concepito come il primo passo per la costruzione del *Brain World*:

A World Encyclopaedia no longer presents itself to a modern imagination as a row of volumes printed and published once for all, but as a sort of mental clearing house for the mind, a depot where knowledge and ideas are received, sorted, summarised, digested, clarified and compared [...]. This Encyclopedic organisation need not be concentrated now in one place; it might have the form of a network. It would centralise mentally but perhaps non physically. Quite possibly it might to a large extent be duplicated. It is its files and its conference rooms which would be the core of its being, the essential Encyclopedia. It would constitute the material beginning of a real World Brain³⁰.

Secondo Wells, la sperimentazione avviata alla British Library della microfilmatura di 4000 libri antichi, era un esempio concreto della fattibilità del suo progetto; si diceva poi certo che in un futuro non lontano avremmo avuto:

²⁷ S. Vaidhyanathan, *La grande G. Come Google domina il mondo e perché dovremmo preoccuparci*, Milano, Rizzoli, 2012, pp. 175-203. Nel 2013 Google ha promosso un nuovo progetto che ricalca il modello Google Books: l'*Open Gallery*, che ospita online collezioni di musei, foto e materiale documentario provenienti da istituzioni o singoli individui disposti a collaborare con il Google Cultural Institute, come nel caso degli accordi stipulati dall'Archivio Alinari e dalla Biennale di Venezia nel 2015.

²⁸ E. Bonacini, *Google e il patrimonio culturale italiano*, in «SCIRES-it», IV (2014)1, pp. 25-40 (<<http://dx.doi.org/10.2423/i22394303v4n1p25>>).

²⁹ Cfr. <<http://www.worldbrainthefilm.com/>>.

³⁰ H.G. Wells, *The Brain Organization of the Modern World. Lecture delivered in America, October and November, 1937*, in Id., *World Brain* (1938), a cura di A.J. Mayne, London, Adamantine Press, 1994, p. 112.

microscopic libraries of record, in which a photograph of every important book and document in the world will be stowed away and [...] any student, in any part of world, will be able to sit with his projector in his own study at his or her convenience to examine any book, any document, in an exact replica³¹.

Nel 1962 Arthur C. Clarke riprese in chiave informatica le idee di Wells sul “cervello mondiale” e ne prevede la realizzazione in due fasi. Entro il 2000, nei paesi sviluppati si sarebbe realizzata la costruzione della World Library accessibile a chiunque su computer, mentre in un secondo tempo la «biblioteca universale» sarebbe stata incorporata in un supercomputer dotato di una superiore intelligenza artificiale in grado entro il 2100 di interagire con l'uomo collaborando alla risoluzione dei problemi del mondo³².

Senza dubbio un filo rosso unisce il *World Brain* all'attuale Rete che molti teorizzano essere il luogo d'elezione dell'intelligenza artificiale che Clarke aveva immaginato realizzata dai supercomputer³³. Va però sottolineato che il cervello mondiale, secondo Wells, sarebbe stato sottoposto al rigido controllo delle élites intellettuali e politiche dell'auspicato futuro *Federal World State*, mentre *Google* è un soggetto privato che può decidere in totale autonomia quali progetti finanziare. In questa ottica anche i milioni di libri digitalizzati sono qualcosa in più di un serbatoio di informazioni e metadati da indicizzare ad uso dei lettori umani. Come dai dati sensibili forniti attivamente da un miliardo di utenti al suo motore di ricerca *Google* estrae valore economico orientando i desideri dei consumatori, così *Google Books* estrae dal patrimonio librario il capitale di conoscenze accumulato dall'umanità da investire nel sogno di trasformare l'azienda nel *World Brain* del futuro.

Una conferma indiretta ci viene da una fonte attendibile, George Dyson, che racconta come, nel 2005, cioè poco dopo la presentazione ufficiale del progetto alla fiera del libro di Francoforte, un ingegnere di *Google* gli avrebbe candidamente confessato che «non stiamo scansionando tutti questi libri per farli leggere ad esseri umani. Li stiamo scansionando per farli leggere a una intelligenza artificiale»³⁴. Con questo obiettivo, nel 2013 Ray Kurzweil è stato messo a capo della divisione *Engineering* di *Google*³⁵. La collaborazione del genia-

³¹ Ivi, p. 112.

³² A.C. Clarke, *Profiles of the Future. An Inquiry into the Limits of the Possible*, New York, Harper & Row, 1962.

³³ Come afferma Ray Kurzweil (*La Singolarità è vicina*, Milano, Maggioli, 2008 p. 255): «Man mano che la conoscenza umana migra sul web, le macchine saranno in grado di leggere e sintetizzare tutte le informazioni».

³⁴ G. Dyson, *La cattedrale di Turing. Le origini dell'universo digitale*, Torino, Codice Edizioni, 2012, p. 362.

³⁵ Kurzweil, esponente di spicco della «terza cultura» teorizzata da John Brockmann (*La terza cultura. Oltre la rivoluzione scientifica*, Milano, Garzanti, 1995; *I nuovi umanisti. Perché (e come) l'arte, la politica, la storia e la filosofia devono tener conto delle moderne scoperte scientifiche*, Milano, Garzanti, 2005), è di sicuro un inventore geniale: tra le sue tante invenzioni ricordiamo la *Kurzweil Reading Machine* e la *Kurzweil Voice Med* per la conversione testo-voce e voce- testo, il sintetizzatore musicale *Kurzweil 259* e il software OCR omni-font. È anche un prolifico scrittore e nei suoi

le e visionario ingegnere può essere interpretata come il segnale di una decisa accelerazione dell'impegno dell'azienda di Mountain View nel campo dell'Intelligenza Artificiale³⁶; infatti, la missione assegnata a Kurzweil è tanto semplice quanto impegnativa: mettere la macchina in grado di comprendere le sfumature e le ambiguità del linguaggio naturale senza l'intervento umano. Un esempio concreto del possibile utilizzo di *Google Books* nel campo dell'I.A. ci è offerto dalla traduzione automatica.

Negli anni Cinquanta del secolo scorso, i primi sistemi computerizzati di traduzione erano basati su norme linguistiche, grammatica, sintassi, *thesaurus* e via dicendo. Nonostante l'impegno profuso per creare modelli basati su teorie linguistiche, capaci ad esempio di distinguere tra un verbo e un sostantivo, i risultati furono deludenti; negli anni novanta si passò dunque ad un diverso

libri di tanto in tanto concede all'affaticato lettore un momento di pausa offrendo dotti riferimenti tratti da autori classici antichi e moderni. Ma evidentemente da buon ingegnere ritiene che la conoscenza del passato non necessiti dello stesso grado di precisione che serve per inventare nuove macchine. Così dalla sua ultima fatica letteraria apprendiamo che «come scriveva nel terzo secolo a.C. il filosofo greco Diogene Laerzio, non ci si può bagnare due volte nello stesso fiume» (*Come costruire una mente. I segreti del pensiero umano*, Milano, Apogeo, 2013, p. 207). Confondere Eraclito con Diogene Laerzio fa venire in mente il latinorum con il quale Don Abbondio cercava di confondere il povero Renzo. Ma se l'errore di citazione del filosofo presocratico può essere considerato un peccato veniale chissà cosa avranno pensato i ricercatori impegnati come Kurzweil nel campo dell'Intelligenza Artificiale, nell'apprendere che al loro *Alivater* Alain Turing «viene attribuito il merito di aver posto gli Alleati nella condizione di poter superare il vantaggio di 3 a 1 di cui godeva la Luftwaffe tedesca rispetto alla Royal Air Force britannica, consentendo loro di vincere la fondamentale Battaglia d'Inghilterra» (come scrive Kurzweil nell'introduzione alla recente ristampa di J von Neumann, *Computer e cervello*, Milano, Il Saggiatore, 2014, pp. 17-18). In realtà non solo gli appassionati di computer sanno che la missione di Turing e dei suoi colleghi era quella di decifrare i codici di crittografia di Enigma, la macchina utilizzata dai tedeschi per le comunicazioni top secret, in particolare con i sommergibili U-Boot. Colossus, la macchina di crittoanalisi ideata da Turing e altri riuscì a porre fine alla superiorità tedesca nell'Atlantico e nei mari del Nord e non sul cielo d'Inghilterra. Ma tant'è, forse Kurzweil considera la storia dell'informatica un hobby per scienziati in pensione. Ovviamente non è difficile trovare giustificazioni per gli svarioni di Kurzweil in filosofia e storia ma la cosa si fa più preoccupante quanto si tratta di questioni strettamente scientifiche che riguardano da vicino i progetti del nostro ingegnere impegnato a liberare gli uomini dal loro passato non solo culturale ma anche biologico. A conclusione di *Come creare una mente* Kurzweil, parlando del limite inesorabile che per il momento la velocità della luce oppone alla colonizzazione dello spazio grazie al *computronium* (così definisce l'unione di materia ed energia che a parità di peso sarà enormemente più potente del cervello umano), con il consueto ottimismo dichiara: «Se esistono modi anche molto raffinati per aggirare questo limite, la nostra intelligenza e la nostra tecnologia saranno abbastanza potenti da sfruttarli. Questo è uno dei motivi per cui la possibilità che i muoni che recentemente hanno percorso i 730 chilometri dall'acceleratore del CERN al confine fra Svizzera e Francia al Laboratorio del Gran Sasso in Italia»... (p. 240) e qui ci fermiamo e stendiamo un pietoso velo sui poveri muoni che come tutti sanno non si sono mai mossi dalla Svizzera.

³⁶ Sempre nel 2013 Google, in collaborazione con l'università di Harvard (Ma) e con il MIT, ha annunciato il progetto Automatic Statistician che ha il compito di superare i problemi degli attuali metodi di *machine learning* che prevedono ancora un notevole intervento umano e nel 2015 ha reso *open source* la piattaforma TensorFlow invitando programmatori e ricercatori a collaborare attivamente all'I.A. di Google in cambio di migliori servizi.

approccio basato sulla raccolta di enormi quantità di testi di una determinata lingua e delle relative traduzioni in un'altra, per poi lasciare alla macchina il compito di indicare il modello probabilistico di traduzione che funziona meglio. Quando su *Google* clicchiamo su «traduci questa pagina», il programma, attualmente il migliore disponibile, ci fornisce la traduzione più probabile. L'obiettivo finale è la “traduzione universale”, affrontata alla maniera di *Google*, senza dizionari, grammatiche o esperti linguisti, ma con il bulimico frullato algoritmico di miliardi di parole. Agli informatici che lavorano a *Google Translate* non è richiesta la conoscenza di nessuna lingua in particolare né specifiche competenze in una qualche disciplina. Ma l'idea che gli algoritmi per la traduzione automatica producano vere e proprie teorie del linguaggio migliori di quelle umane è affascinante per chi ritiene che il cyberspazio possieda una sorta di auto-conoscenza che la macchina può esplorare alla ricerca di *pattern* linguistici nascosti immediatamente utilizzabili senza il ricorso a teorie che ne spieghino l'origine. Come ha sottolineato Michel Nielsen,

[poiché] questi modelli per la traduzione sono in un certo senso delle teorie o delle spiegazioni della traduzione, dovremmo quantomeno considerare seriamente l'ipotesi che tali modelli statistici esprimano delle verità che le spiegazioni più convenzionali della traduzione linguistica non possiedono³⁷.

È legittimo sospettare che *Google* faccia un ampio ricorso ad esperti, ma l'idea stessa che un computer possa imparare a tradurre desta perplessità nei letterati e nei lettori che sanno quanto è complessa l'arte del tradurre un brano di prosa o una poesia, e che non tutti sono disposti ad affidare ad un algoritmo la scelta se la nave di Ulisse solcava il «divido mare» o il «mare color del vino».

Anche se al livello attuale gli algoritmi probabilistici applicati a *Google Books* non possono competere con la superiore qualità della comprensione umana della complessità del linguaggio naturale, già oggi alcuni ricercatori considerano le risorse offerte da *Google* una forma embrionale della *knowledge machine* del futuro. Come hanno scritto di recente autorevoli ricercatori californiani impegnati nel campo delle *Digital Humanities*,

è giunto il momento di considerare il potenziale della lettura della macchina (*machine reading*), un'analisi che sfrutta metodi computazionali per individuare tendenze, relazioni e correlazioni testuali. Il ricorso a strumenti e metodi digitali s'impone nel momento in cui la produzione di informazione supera di gran lunga le facoltà di comprensione umane. Il modo migliore per leggere i libri, oggi, è di non leggerli. O meglio, chiedere alle macchine di leggerli per noi³⁸.

In effetti, se ci limitiamo alla quantità, non si può non dare ragione a quanti praticano la «lettura a distanza» teorizzata da Franco Moretti, docente di lette-

³⁷ M. Nielsen, *Le nuove vie della scoperta scientifica. Come l'intelligenza collettiva sta cambiando la scienza*, Torino, Einaudi, 2012, p. 145.

³⁸ A. Burdick *et al.*, *Umanistica_Digitale*, Milano, Mondadori, 2014, p. 64.

ratura a Stanford³⁹. Proviamo ad immaginare un ipotetico studioso che legge due libri a settimana per 60 anni ai quali aggiunge ogni anno 100 articoli di riviste specializzate e 200 opere consultate a vario titolo. Il risultato finale di una vita di studi sarà quantitativamente irrisorio se paragonato alle tecniche automatiche di analisi testuale che già oggi è possibile effettuare tra i milioni di libri digitalizzati da *Google Books*. Nel 2010, con il supporto finanziario di *Google*, Erez Liberman-Aiden e Jean-Baptiste Michel hanno creato un software per l'analisi quantitativa di circa cinque milioni di libri. La versione del programma, *Ngram Viewer*⁴⁰, messa a disposizione degli utenti gratuitamente da *Google*, consente la visualizzazione grafica dei risultati della ricerca effettuata su singole parole o frasi. Secondo gli autori, nell'era dei *Big Data* l'analisi automatica dei testi rappresenta una vera e propria rivoluzione copernicana degli studi umanistici e delle scienze sociali. I primi risultati della ricerca sono stati presentati nel libro *Uncharted*, in cui gli autori definiscono «culturomica» questo nuovo approccio alla cultura e ai comportamenti umani⁴¹. La nuova disciplina si pone l'obiettivo di mappare l'intero patrimonio culturale dell'umanità così come ha fatto la genomica per quello genetico⁴².

Anche se la mia personale esperienza con il programma è stata limitata, l'impressione che ne ho ricavato è che gli autori sopravvalutino l'utilità dell'analisi probabilistica delle tendenze culturali, di idee e interessi nel corso dei decenni o secoli. In molti casi i dati empirici generati dalla "lettura a distanza" aggiungono poco al già noto; per fare un esempio, gli storici considerano la censura da parte dei nazisti dell'arte "degenerata" o la cancellazione di Trockij dalla memoria collettiva dell'Unione Sovietica di Stalin come fatti ormai acquisiti. In particolare, nel campo della critica letteraria i risultati della ricerca possono persino produrre un'immagine capovolta del valore di un autore, la cui fecondità artistica non è quantificabile. Ad esempio, dovremmo cambiare il giudizio sull'importanza di Charles Dickens o di Herman Melville solo perché i risultati della macroanalisi condotta da Matthew I. Jockers di 3592 romanzi pubblicati tra il 1780 e il 1900 indica che lo stile e i temi dei romanzi di Jane Austen e Walter Scott hanno influenzato un numero maggiore di autori rispetto a Dickens e Melville? Si può valutare il valore di un'opera letteraria utilizzando un algoritmo che misura l'influenza di un autore e i legami tra i romanzi con criteri analoghi a quelli utilizzati da *Page Rank* per classificare la rilevanza delle pagine Web?⁴³

³⁹ F. Moretti, *Distant Reading*, New York, Verso Books, 2013; Id., *Graphs, Maps, Trees: Abstract Models for Literary History*, New York, Verso Books, 2007.

⁴⁰ N-grammi nella terminologia della linguistica computazionale sono le occorrenze di una parola o di una frase misurate nel corso di un certo periodo di tempo.

⁴¹ E. Aiden - J.B. Michel, *Uncharted, Big Data as a Lens on Human Culture*, New York, Riverhead Books, 2013.

⁴² Secondo l'etologo ultra-darwinista Richard Dawkins il gene, ha il suo omologo celebrale definito mene che costituirebbe l'unità di informazione elementare dell'evoluzione culturale soggetta a meccanismi di selezione analoghi a quella biologici.

⁴³ M.L. Jockers, *Macroanalysis: Digital Methods and Literary History (Topics in the Digital Humanities)*, Champaign (Il), University of Illinois Press, 2013.

Nel passaggio dalla “lettura diretta” e scrupolosa di testi selezionati alla “lettura a distanza” di milioni di testi digitali molto si perde, perché, secondo il detto attribuito ad Albert Einstein, «non tutto quel che conta può essere contato e non tutto quello che può essere contato conta». Ma di certo non basta un aforisma (oggi diremmo un *tweet*) per incrinare la fiducia nelle magnifiche sorti e progressive della tecnoscienza nell’era dei *Big Data*. Soprattutto per quanti confidano che “solo *Google* ci salverà” dal diluvio di dati che ci sommerge.



Roberto Mazzola
ISPF-CNR, Napoli
mazzola@ispf.cnr.it

– Google books e le scienze (post)umane

Citation standard:

MAZZOLA, Roberto. Google Books e le scienze (post)umane. Laboratorio dell'ISPF. 2015, vol. XII. DOI: 10.12862/ispf15L405.

Online: 30.12.2015

ABSTRACT

Google books and the (post)Humanities. This article examines some steps of the evolution of the relationship between computer science and humanities and focuses on the activity of Google in the digitization, preservation and dissemination of cultural assets. In particular it critically analyzes the role of Google Books and the meaning of the Big Data in the Digital Humanities.

KEYWORDS

Google books; Digital humanities; Big Data; ICT and Humanities

SOMMARIO

L'articolo ripercorre alcune tappe dell'evoluzione del rapporto tra informatica e scienze umane e si sofferma sull'attività di Google nella digitalizzazione, conservazione e diffusione dei beni culturali. In particolare si analizza in modo critico il ruolo di Google Books e il significato dei Big Data nel campo delle Digital Humanities.

PAROLE CHIAVE

Google books; Digital humanities; Big Data; Informatica e scienze umane

