

8



STRUMENTI

Atti dei seminari di studio
Pisa, Scuola Normale Superiore
10 marzo-12 dicembre 2005

Seminari Signum 2005

a cura
di Simonetta Bassi

con la collaborazione di
Francesca Di Dio



EDIZIONI
DELLA
NORMALE

Indice

Presentazione	7
Il progetto Fondi ed Archivi Digitali: la tecnologia al servizio degli Istituti culturali ANDREA BOZZI, MARCO PASSAROTTI	9
Tecnologie Grid e calcolo autonomico ANDREA RIPOLI	35
La magia della scrittura: modelli neurolinguistici nel messaggio scritto ALESSANDRO LUCCHINI	45
Apprendimento automatico di ontologie da testi: selezione di relazioni semantiche da un <i>corpus</i> di biologia molecolare MASSIMILIANO CIARAMITA	59
Etica di Internet ADRIANO FABRIS	83
‘Archiviare’ il Web: iniziative e progetti STEFANO VITALI	99
Paesaggi di Informazione MASSIMO BERGAMASCO, EMANUELE RUFFALDI, CHIARA EVANGELISTA	129
Combinazione di parole e spazi semantici: un’analisi computazionale dei testi di Giordano Bruno ALESSANDRO LENCI, FELICE DELL’ORLETTA	139
Progettare oggi: dalle forme meccanicistiche alle forme digitali alle forme geneticamente mutate, nella continua ricerca della bellezza PAOLO DEGANELLO	159

Presentazione

Gli articoli contenuti in questo volume sono il risultato del lavoro seminariale svolto presso Signum, Centro di ricerche informatiche per le discipline umanistiche della Scuola Normale Superiore di Pisa, diretto da Michele Ciliberto. Il laboratorio ha organizzato questi incontri per favorire la diffusione della conoscenza teorica e dei progetti nell'ambito della ricerca informatica applicata agli studi umanistici. Ma non solo: Signum, infatti, si occupa di studiare, progettare e sviluppare soluzioni algoritmiche per l'analisi, l'archiviazione, la catalogazione e la ricerca di sequenze, intendendo con tale termine una qualunque stringa di simboli significativa: un testo o una pagina web, un file XML o un record archivistico (o anche una sequenza genomica). La ricerca di base viene applicata alla realizzazione di prodotti software per la gestione di archivi e catalogazione di documenti (sia strutturati che semistrutturati); per la compressione dei dati, per la realizzazione di motori di ricerca e di interfacce grafiche innovative per l'utilizzo del Web. Questa attività comporta un continuo dialogo con discipline diverse (la linguistica computazionale, la bioinformatica...) e contestualmente favorisce l'emergere di problematiche molteplici, di natura etica (l'accessibilità e la condivisione degli strumenti), comunicativa (le forme di scrittura del Web), progettuale (la grafica in rete) e cognitiva (l'organizzazione della conoscenza): il lavoro del laboratorio si trasforma allora in uno spazio aperto in cui agiscono molteplici sollecitazioni cui devono rispondere in modo integrato umanisti e informatici, tecnici e filosofi. I saggi contenuti in questo volume rappresentano un primo sondaggio in quello spazio che per molti versi rimane ancora da costruire.

Il progetto Fondi ed Archivi Digitali: la tecnologia al servizio degli Istituti culturali

1 Introduzione

Il progetto Fondi ed Archivi Digitali (FAD) si colloca nel quadro delle attività di aggiornamento tecnologico in atto presso gli istituti culturali che intendono valorizzare il proprio patrimonio bibliotecario ed archivistico anche grazie a impegnative campagne di digitalizzazione. Esso è stato finanziato dal Ministero per i Beni e le Attività Culturali nell'ambito della specifica legge «Interventi straordinari nel settore dei beni e delle attività culturali»¹. La proposta, presentata dalla Fondazione Conti di Fiesole, ha coinvolto altri tre istituti culturali (il Gabinetto Vieusseux di Firenze; l'Istituto Papirologico Vitelli di Firenze; la Fondazione Roselli di Torino) ed un istituto di ricerca del CNR (Istituto di Linguistica Computazionale di Pisa)².

Il sistema informatico che è alla base delle funzionalità di FAD³ rappresenta l'evoluzione di un prototipo, ideato presso il CNR, per lo studio e l'edizione elettronica di documenti digitali (manoscritti e a stampa) e sviluppato in più occasioni successive. Fra queste, la prima e più significativa è costituita da BAMBI (*Better Access to Manuscripts and Browsing of*

¹ Marco Passarotti è autore della sezione 5.

² Legge 21/12/1999, n. 513, nota col nome di «Legge Biscardi».

³ La proposta è stata redatta da Andrea Bozzi; il presente contributo si basa in parte sulla relazione conclusiva inviata al Ministero, contenente i risultati scientifici e tecnologici ottenuti in parte sul manuale per gli utenti del sistema FAD, curato da Marco Passarotti. Per gli aspetti di carattere metodologico che sono alla base dell'applicazione e per i relativi riferimenti bibliografici, si rinvia a A. BOZZI, *Edizione elettronica e filologia computazionale*, in A. STUSSI, *Fondamenti di critica testuale*, Bologna, Il Mulino 1994.

³ Lo sviluppo del *software* è stato realizzato da META. s.r.l. di Lucca, che si è aggiudicata l'appalto.

Images), progetto approvato e cofinanziato dalla Commissione Europea nel settore delle applicazioni telematiche per le biblioteche, nell'ambito del III Programma Quadro di Ricerca e Sviluppo⁴.

Gli obiettivi principali di FAD sono stati i seguenti:

- consentire una più semplice fase di archiviazione, gestione e consultazione di materiali librari e di archivio attualmente non accessibili né catalogati o solo in parte catalogati;
- realizzare una conversione ottimale degli originali su supporto numerico ed una gestione informatizzata degli archivi digitali per facilitarne conservazione ed utilizzo;
- promuovere la visibilità internazionale di documenti librari e di archivio di particolare valore storico e culturale, attualmente conosciuti solo da un ristretto numero di specialisti;
- realizzare strumenti innovativi per produrre l'edizione elettronica (trascrizione, studio, indicizzazione, ecc.) dei documenti digitali, strumenti che, pertanto, possono configurarsi come nuovi servizi per le biblioteche e gli archivi digitali;
- creare risorse documentali digitali fruibili in rete mediante metodologie evolute, basate su Intranet (servizi FAD in ambito locale) e Internet, tenendo presenti sia la protezione di documenti coperti da *copyright*, sia le indicazioni tecniche relative alla BDI (Biblioteca Digitale Italiana)⁵;
- mettere a punto una piattaforma generalizzata (*server* FAD) applicabile sia in ambito locale, per offrire servizi specializzati (per esempio, il lavoro cooperativo prodotto da più utenti contemporaneamente su uno stesso *corpus* di documenti), sia in ambito Internet, per la consultazione degli archivi e delle informazioni che vi sono state associate da parte di specialisti che abbiano manifestato la propria disponibilità a rendere visibile ad altri i risultati del proprio lavoro di analisi e studio.

2 Aspetti tecnologici: la piattaforma informatica

⁴ Per maggiori informazioni si veda *Better Access to Manuscripts and Browsing of Images. Aims and Results of an European Research Project in the Field of Digital Libraries*, a cura di A. Bozzi, Bologna, CLUEB 1997. Il volume miscelaneo descrive le fasi di progettazione e di sviluppo dei componenti realizzati sulla base di codici manoscritti di epoca medievale in lingua latina ed italiana.

⁵ Si veda il sito <<http://www.iccu.sbn.it/bdi.html>>.

Per la realizzazione degli obiettivi indicati, è stato necessario progettare una piattaforma informatica (*hardware e software*) con funzionalità tali da consentire la produzione degli archivi coerente coi sistemi di catalogazione già adottati dagli istituti partecipanti, e il raccordo fra archivi già disponibili in formato elettronico, ma realizzati con sistemi diversi. Si è considerato opportuno progettare una struttura a due livelli: un livello locale (piattaforma interna ad ogni singola istituzione, principalmente dedicata alla produzione) e un livello centralizzato (piattaforma esclusivamente dedicata alla consultazione).

Per quanto riguarda il primo livello (rete locale), un *server* è destinato all'immagazzinamento dei dati (per esempio, i cataloghi, l'archivio delle immagini ad alta risoluzione), alla gestione degli accessi e all'interscambio dei dati con il *server* centrale. Il *server* locale è collegato a un certo numero di postazioni (*client*) che possono essere sia di consultazione sia di lavoro, a seconda delle caratteristiche tecniche (memoria di elaborazione, spazio disco, ecc.).

Il secondo livello è principalmente costituito da un *server* centrale connesso in modo permanente ad Internet attraverso una linea dedicata ed in grado di fornire le informazioni sui dati degli archivi di tutti gli istituti partecipanti al progetto. All'interno di esso sono presenti le informazioni complete di tutti i cataloghi e, qualora disponibili, anche le immagini a bassa risoluzione dei documenti digitalizzati. Il *server* centrale offre una serie di interfacce opportunamente disegnate per accedere alle diverse tipologie di documenti e per interrogarli. Bisogna segnalare, infatti, che, come si vede dagli esempi riportati nella seconda parte del contributo, la documentazione presente in FAD riguarda dati di varia natura: documentazione archivistica e libraria redatta su carta, frammenti di papiro in lingua greca, oggetti museali e reperti archeologici. Viene fornita, inoltre, la possibilità di effettuare operazioni di ricerca di documenti appartenenti ad archivi di istituzioni diverse come se si trattasse di un unico *corpus*.

I dati presenti nel *server* centrale sono aggiornati da appositi moduli che operano in ciascuno dei sottosistemi locali attraverso un collegamento Internet di tipo ISDN⁶ o superiore, effettuato mediante una

⁶ L'acronimo ISDN (*Integrated Services Digital Network*) fu adottato per la prima volta nel 1971 dal IX gruppo di studio del CCITT (*International Telephone and Telegraph Consultative Committee*), ora conosciuto come ITU (*International Telecommunications Union*), che ha lo scopo di coordinare e standardizzare le telecomunicazioni internazionali. In Italia,

connessione sicura (VPN, *Virtual Private Network*), grazie alla quale i dati vengono crittografati⁷.

3 Funzionalità

La piattaforma informatica realizzata è composta da un insieme di moduli *software*, ciascuno dei quali svolge un compito specifico nell'ambito del sistema complessivo. L'utente può accedere alle diverse funzionalità del sistema mediante un menù grafico, posto nella parte alta della finestra del *browser*, costituito da un insieme di bottoni: l'azione prodotta quando se ne seleziona uno è intuitiva anche se, per comodità dell'utilizzatore, appare una didascalia concisa (*tool tip*) tutte le volte che il puntatore del *mouse* vi si sovrappone.

3.1 Organizzazione e gestione dei documenti

Il sistema consente di archiviare in uno schedario elettronico i dati

nel 1992, venne costituito il 'Consorzio Promotore ISDN', allo scopo di diffondere la cultura dei nuovi servizi, di stimolare la nascita di applicazioni innovative e di attivare forme concrete per rendere familiari alla clientela le potenzialità offerte dalle strutture della rete ISDN (tra i fondatori del Consorzio vi erano il gestore del servizio pubblico Telecom Italia e molte aziende private costruttrici di prodotti per telecomunicazioni, informatica e *software* applicativo. I servizi attualmente definiti nelle specifiche di ISDN consentono trasmissione in fonia, trasmissione in fonia ad alta qualità, trasmissione di musica, fax, foto e immagini in movimento.

⁷ Una rete VPN permette a computer ubicati in sedi fisiche diverse di stabilire un collegamento tramite Internet. Questa soluzione elimina la necessità di ricorrere a costose linee dedicate poiché la connessione a Internet permette di collegare sia sedi diverse sia utenti remoti. Poiché le connessioni a Internet sono connessioni pubbliche, e quindi non protette per definizione, sono esposte al rischio che i pirati informatici possano intercettare e modificare i dati trasmessi sul Web. Con una rete VPN è tuttavia possibile crittografare i dati e inviarli solo ad un computer (o gruppo di computer) specifici o, in altre parole, di creare una rete privata che è accessibile solo agli utenti autorizzati. La rete in questione è però una rete virtuale poiché il collegamento tra i computer remoti non è fisico ma basato sul Web: per poter utilizzare un'applicazione installata su un sistema centrale, gli utenti esterni che lavorano su *client* fuori sede devono semplicemente collegarsi allo specifico sito web e immettere una password.

relativi al documento originale o alla copia fotografica utilizzata per il processo di digitalizzazione: i documenti sono catalogati con metadati descrittivi secondo una prassi ormai consolidata. La struttura delle schede di catalogazione è conforme alle richieste degli Istituti Culturali e la creazione degli archivi è facilitata da un'interfaccia grafica semplice ed intuitiva. Il lavoro di compilazione viene assistito, per esempio, fornendo all'operatore liste di termini stabilite a priori (liste chiuse) o incrementabili (liste aperte), oppure facilitando la distinzione fra i campi che possono accogliere un valore unico e quelli che, invece, possono riceverne più di uno. È possibile, inoltre, precompilare alcuni campi in modo che i valori vengano inseriti automaticamente su gruppi di schede omogenee, con un notevole risparmio di tempo e una diminuzione del rischio di inserire errori da parte dell'operatore. Un opportuno promemoria facilita il ritrovamento dell'ultima scheda catalogata e, nel caso di catalogazione incompleta, dell'ultimo campo riempito. Per quanto riguarda aspetti di amministrazione, ogni operatore viene registrato e possiede il diritto di modificare successivamente i dati che ha immessi.

Le informazioni delle schede catalografiche possono essere interamente esportate nei formati XML e PDF: questa seconda opzione rende particolarmente semplice l'operazione di stampa che, a scelta dell'utente, mostra i dati compilati di ogni singola scheda in forma compatta su un foglio di dimensioni A4.

I dati sono memorizzati sul *server* locale all'interno di una base di dati relazionale. La consultazione si attua col metodo ormai diffuso delle maschere di ricerca semplice o avanzata: si possono usare operatori booleani e liste di termini predisposte dal sistema sulla base dei dati già registrati.

Le diverse tipologie di documenti scelti dagli Istituti culturali per il progetto FAD hanno reso necessaria la progettazione e realizzazione di archivi differenti, ma nello stesso tempo è stato indispensabile individuare il maggior numero di elementi di omogeneità fra essi al fine di favorire la ricerca e il recupero di dati che avessero caratteristiche comuni⁸. La diversità degli archivi si riflette evidentemente anche sulla

⁸ Si consideri che, per esempio, l'Istituto Papirologico Vitelli ha catalogato anche alcune fatture relative all'acquisto di materiali utili agli scavi archeologici, e grazie ad esse è possibile avere dettagli sugli strumenti del mestiere archeologico impiegati nei primi decenni del '900, mentre il Gabinetto Vieuxseux ha schedato e digitalizzato libri stam-

struttura delle schede catalografiche che hanno un insieme di campi variabile da qualche decina (Istituto Papirologico Vitelli) a circa un centinaio (Fondazione Conti), suddivisi per pagine e, quando necessario, separati in sezioni.

Sono stati infine prodotti moduli per l'importazione automatica nell'archivio FAD dei dati già parzialmente memorizzati presso alcuni Istituti con sistemi di catalogazione in alcuni casi superati, non più aggiornati o usciti dal mercato. Ciascun modulo realizza tale integrazione in modo automatico o semiautomatico, richiedendo eventuali interventi di verifica e correzione da parte di un operatore autorizzato (amministratore).



Fig. 1. Esempio di scheda catalografica (Gabinetto Vieusseux)

3.2 Digitalizzazione e ritocco di immagini

Una postazione di lavoro presente nella rete interna di ciascuna Istituzione è abilitata ad eseguire la conversione dei documenti in formato digitale ed è, per tale motivo, collegata ad uno *scanner*. È possibile acquisire direttamente le immagini da catalogare all'interno del sistema, senza dover ricorrere a programmi esterni; le immagini digitali vengono

pati alla fine del XIX sec., con l'intento di studiare le illustrazioni relative alla letteratura di viaggio. Questi dati sono certamente di tipo archivistico il primo e librario il secondo, ma esiste un legame fra i due, se non altro di tipo cronologico e tematico (l'archeologia classica, i monumenti e le memorie storiche, ecc.). Ciò ha spinto a trovarne delle relazioni già in fase di schedatura; su questo aspetto, tuttavia, la sperimentazione fino ad oggi eseguita necessita di affinamenti e conferme.

salvate sul disco del *server* locale in maniera indipendente dall'operatore. Egli, infatti, può assegnare un nome ed una posizione progressiva al documento, dati che verranno poi utilizzati quando si chiederà al sistema di portarlo in visione, mentre il nome del file e la posizione su disco vengono decise ed assegnate dal *software* del *server* locale. I primi cioè (nome e posizione del documento assegnata dall'operatore) non influiscono sulla memorizzazione fisica (il cosiddetto 'salvataggio') dei secondi: anche questa soluzione tecnica consente di evitare errori che potrebbero essere numerosi soprattutto nel caso di grandi volumi di dati da convertire in formato digitale⁹. L'immagine acquisita può tuttavia essere salvata anche sulla postazione di lavoro (*client*), affinché risulti più agevole modificarne le caratteristiche grafiche (per esempio, la luminosità o il contrasto) mediante un programma di fotoritocco del quale l'operatore sia già esperto, ed essere inviata al *server* in un momento successivo. È possibile, infine, eseguire anche l'operazione inversa, ovvero riportare le immagini dal *server* al *client*, una volta che, naturalmente, vi siano state memorizzate.

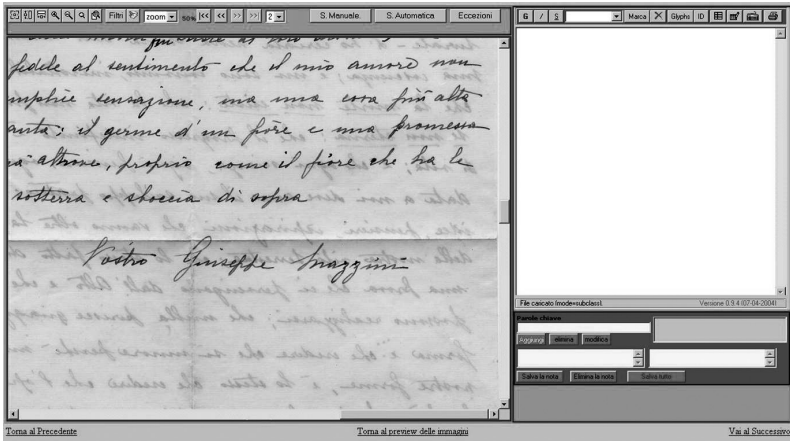
Uno specifico programma attivo sul *server* analizza tutte le immagini e ne produce copie a bassa risoluzione: esse, da un lato, agevolano la consultazione veloce degli archivi digitali soprattutto quando hanno raggiunto dimensioni molto consistenti, dall'altro diventano patrimonio del *server* centrale FAD nel quale, per ragioni di riservatezza e di salvaguardia dei diritti di riproduzione, non sono ospitate, se non su esplicita autorizzazione, le immagini corrispondenti ad alta risoluzione.

Le operazioni di fotoritocco, oltre a quanto visto in precedenza, possono anche essere eseguite con una serie di filtri che FAD mette a disposizione del catalogatore-digitalizzatore: la fatale discontinuità qualitativa delle immagini (disomogeneità delle tonalità cromatiche o dei livelli del grigio, disomogeneità dei contrasti, ecc.) comporta, infatti, la necessità di effettuare un ampio spettro di elaborazioni (luminosità/contrasto, *sharpen*, *edge detection*, gestione cromatica, focus, separazione per canali, rotazioni). Queste funzioni, se ben adoperate, possono agevolare l'interpretazione di parti danneggiate e illeggibili di un documento: questo caso si è verificato, per esempio, per rendere visibili le date dei timbri postali sulle buste, per leggere il testo scritto su carta velina (come nel caso

⁹ Abbiamo effettivamente constatato che l'operatore introduce numerosi errori, specialmente quando il documento da digitalizzare è costituito da molte unità (fogli di uno stesso faldone, pagine di uno stesso libro, fotografie di uno stesso album, ecc.).

dell'esempio in fig. 2) e, come è facilmente comprensibile, per trascrivere il testo conservato sui frammenti dei papiri greci posseduti dall'Istituto Vitelli (si veda un esempio alla fig. 6).

Fig. 2. Risultato di *image enhancement* per facilitare la trascrizione del testo

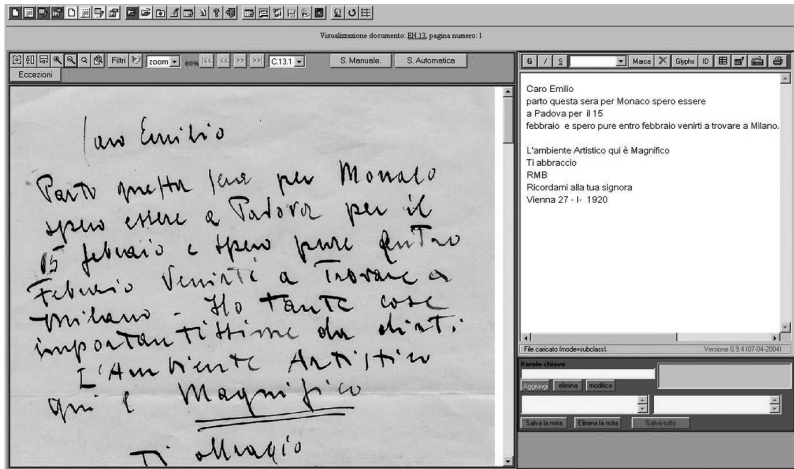


3.3 Leggio-scrittoio digitale

Per quanto riguarda la ricerca e la visualizzazione dei dati, oltre alle normali possibilità di selezione (per autore, titolo, soggetto, lingua), l'archivio accetta selezioni personalizzate e selezioni dipendenti dal particolare archivio su cui la ricerca viene effettuata. Le pagine dei documenti trovati sono rappresentate in forma miniaturizzata e accompagnate da un'etichetta con il relativo riferimento numerico (per esempio, il numero della pagina nel caso di un libro). La scelta di un documento da parte di chi effettua la consultazione comporta la visualizzazione di tutte le miniature delle pagine, che sono presentate a gruppi di 8 per volta. La selezione di una qualsiasi di esse consente di portare in visione quella corrispondente in dimensioni tali da entrare all'interno di un leggio-scrittoio digitale. Qui si trovano: nella parte sinistra, il riquadro contenente l'immagine; nella parte destra, l'eventuale trascrizione e, in basso, alcune finestre entro le quali si possono scrivere annotazioni e commenti. Nel caso in cui un documento aperto sia costituito da più pagine, è possibile accedere alla pagina precedente o alla successiva mediante l'utilizzo di frecce rivolte, rispettivamente, verso sinistra o verso destra. Un insieme di strumenti permette, inoltre, una migliore fruizione dell'immagine: zoom di diverse tipologie (graduale, visione completa nel senso della larghezza, visione completa nel senso dell'altezza, *zoom box* per selezionare ed ingrandire solo una parte, ecc.) e, come detto pre-

cedentemente, filtri per manipolazioni di carattere grafico-cromatico.

Fig. 3. L'ambiente di lavoro principale: immagine e testo trascritto



3.4 Trascrizione e interscambio dei documenti

L'immagine di un documento contenente testo o parti di testo, una volta aperto e visualizzato sul leggio-scrittoio, può essere trascritto con l'utilizzo di un semplice programma di videoscrittura dotato delle funzioni di inserimento, selezione, copia, incolla, cancellazione. Esso permette la stesura di testi in formato XML¹⁰, senza la necessità di inserire manualmente gli elementi di marcatura (i cosiddetti *tag*) necessari per caratterizzare le unità del testo (le parole): una procedura automatica permette, infatti, il riconoscimento dei separatori fra le stringhe-parola. È ovviamente consentito inserire dei marcatori aggiuntivi alle parti di testo

¹⁰ Dal punto di vista dell'evoluzione dei linguaggi di codifica delle informazioni testuali, al fine di superare i limiti imposti dall'*HyperText Markup Language* (HTML), il comitato internazionale dello *Standard Generalized Markup Language* (SGML), formatosi nel 1996 sotto gli auspici del *World Wide Web Consortium* (W3C), ha proposto l'*Extensible Markup Language* (XML), un vero e proprio sistema che consente di ottenere un formato universale di pubblicazione dei dati in rete. XML è un sottoinsieme dello SGML e fornisce al progettista la possibilità di ottenere un'effettiva flessibilità dell'informazione favorendo il riutilizzo e lo scambio dei dati con altre applicazioni. Un buon manuale aggiornato di codifica dei testi nel settore dell'informatica umanistica è stato di recente curato da Elena Pierazzo, *La codifica dei testi. Un'introduzione*, Roma, Carocci 2005.

selezionate (singole parole o espressioni), effettuando la scelta fra quelli proposti da un menù a tendina che si apre agendo sul tasto destro del mouse. Per ogni marcatore inserito è possibile, inoltre, immettere un insieme di attributi (proprietà del *tag*). I *tag* e gli attributi utilizzati in FAD costituiscono un sottoinsieme significativo dello standard TEI (*Text Encoding Initiative*)¹¹, oltre ad altri marcatori specifici che sono definiti in differenti *namespace*. La trascrizione del documento può essere visualizzata come testo puro, cioè privo di richiami agli elementi di marcatura (si veda la fig. 3), come testo marcato secondo lo standard XML-TEI, ove i *tag* utilizzati sono visualizzati in modo grafico (si veda la fig. 4), oppure come testo in formato HTML (si veda la fig. 5). I testi sono trascritti con caratteri in formato Unicode. L'Istituto Papirologico Vitelli ha potuto inserire nel leggito-scrittoio le trascrizioni dei testi papiracei in greco antico (si veda la fig. 6) anche per mezzo di una tastiera virtuale che facilita l'immissione di caratteri non standard (caratteri con accenti, con spiriti, con simboli sottoscritti, segni diacritici tipici per esprimere convenzioni editoriali in papirologia, ecc.).

Funzioni di esportazione del semplice testo, del testo annotato, del testo con le corrispondenze con l'immagine (si veda successivo paragrafo 3.6) consentono grande interscambiabilità dei dati linguistici presenti nelle trascrizioni di archivi FAD con altri archivi testuali presenti e consultabili in Internet.

Fig. 4. La codifica XML-TEI del testo della trascrizione

Fig. 5. La codifica HTML del testo

¹¹ Per quanto riguarda le informazioni di tipo testuale, lo standard che si è affermato nell'ultimo decennio e che si è costantemente evoluto è rappresentato dalla TEI che, all'indirizzo Internet <<http://www.tei.org>>, mette a disposizione le linee guida specifiche per chi voglia lavorare nel settore dell'elaborazione elettronica in ambito filologico, linguistico e letterario. Coloro che le adottano hanno la garanzia che il modo col quale essi caratterizzano e codificano le varie componenti del testo non solo è compatibile con i sistemi di accesso e navigazione in Internet, ma risulta anche omogeneo con la maggior parte dei progetti realizzati a livello internazionale nell'*humanities computing* sia da istituzioni universitarie e di ricerca, sia da soggetti privati come, per esempio, molte case editrici.

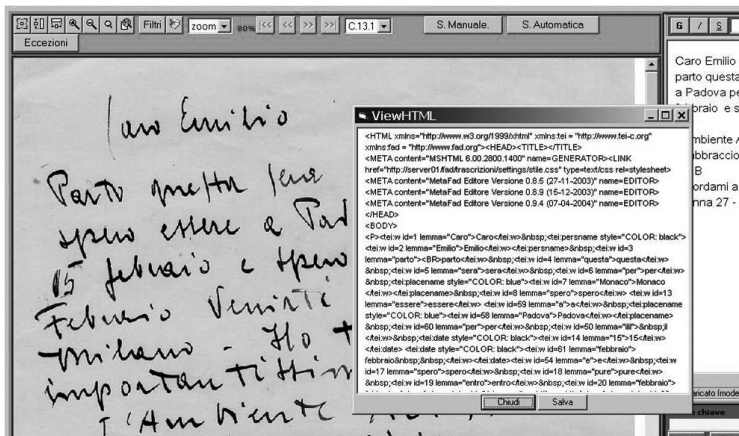
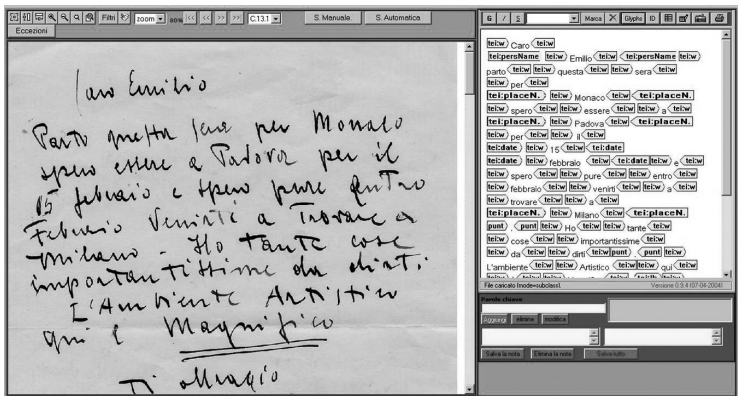
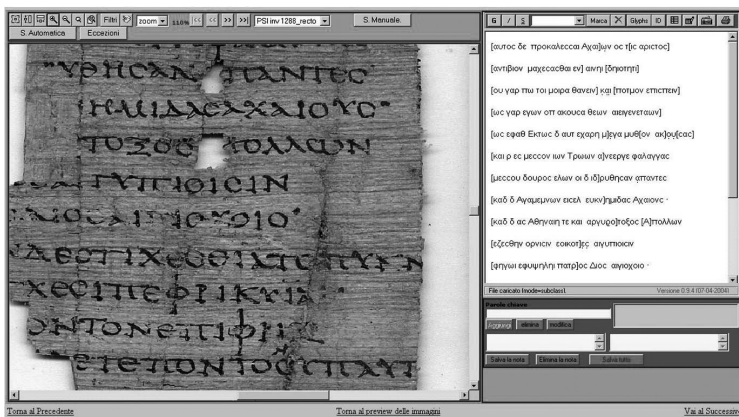


Fig. 6. Gestione di documenti in alfabeto greco antico (Unicode)



Indicizzazione dei testi

Tutte le parole (forme) inserite in un documento durante la trascrizione nonché le informazioni contenute in *tag* specifici e in alcuni loro attributi sono ordinate dal sistema in un insieme di indici nei quali compaiono i riferimenti ai luoghi del testo ove ogni singola unità compare. Nella fase di sperimentazione abbiamo creato indici per le forme, per l'attributo 'lemma' del *tag* 'Word' (con cui è possibile, ad esempio, creare un indice dei lemmi, inserendo manualmente¹² il lemma corrispondente ad una forma), e per alcuni *tag* particolari: i luoghi, le persone, le date, le parole straniere¹³.

3.6 *Corrispondenza testo-immagine*

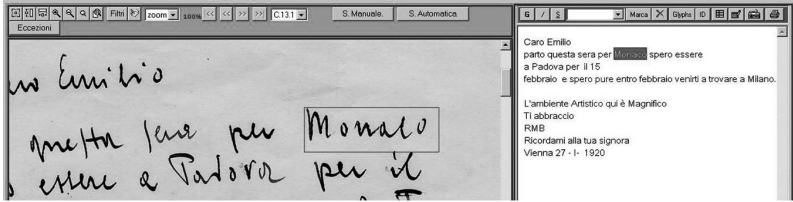
Una caratteristica particolarmente innovativa di FAD è costituita da un componente *software* che realizza la correlazione automatica fra il documento-immagine e il documento-testo: ne consegue che la selezione di una parola qualsiasi sul testo della trascrizione produce contemporaneamente la segnalazione della riga nella quale essa si trova all'interno dell'immagine digitale. Nel caso di immagini di documenti a stampa, dattiloscritti e manoscritti, purché con andamento molto regolare e senza danneggiamenti alla superficie scrittoria, il programma è in grado di operare con maggiore precisione riuscendo a far corrispondere ogni parola del testo con la zona dell'immagine nella quale è inscritta. Un opportuno strumento consente, tuttavia, la correzione manuale degli eventuali errori commessi dal sistema automatico. Lo stesso strumento, che è costituito da un'interfaccia grafica di utilizzo molto semplice e veloce, permette anche di procedere all'inserimento manuale delle corrispondenze fra parole nel testo e nell'immagine, qualora naturalmente si ritenga che questa operazione abbia utilità ai fini interpretativi e qualora la complessità strutturale del documento renda impossibile l'associazione automatica. La visualizzazione della riga o della zona sull'immagine digitale si ottiene anche selezionando uno dei luoghi dell'indice in cui una parola è attestata. Viceversa: la selezione di una parola sull'immagine provoca la immediata visualizzazione della riga o della parola corri-

¹² Il sistema è predisposto per interagire con moduli di analisi linguistica quali, per fare solo un esempio, gli analizzatori morfologici automatici o i sistemi di lemmatizzazione in grado di attribuire l'entrata lessicale alle forme flesse del testo. Naturalmente, casi di omografia devono essere disambiguati manualmente.

¹³ Si veda anche il paragrafo 5, in particolare il commento alla figura 8.

spondente sul testo della trascrizione (si veda la fig. 7)¹⁴.

Fig 7. Corrispondenza automatica tra una parola nel testo e nell'immagine



3.7 Annotazione del testo trascritto

FAD consente all'utilizzatore del leggio-scrittoio di associare annotazioni e commenti ad una parola o a un gruppo consecutivo di parole della trascrizione. Queste note possono essere costituite anche da varianti al testo¹⁵. Il sistema è in grado di gestire annotazioni diverse per utenti differenti e di collegare le annotazioni al testo XML della trascrizione, al fine di permettere esportazioni più articolate.

4 I dati archivistici e bibliotecari in FAD

¹⁴ Questo componente è risultato particolarmente interessante per analisi di tipo paleografico in relazione alla documentazione papiracea. Esso potrebbe avere anche un ruolo didattico in una prospettiva di utilizzo per la formazione a distanza: la concordanza di tutte le forme grafiche con le quali una parola si presenta in uno o più documenti favorisce lo sviluppo di competenze paleografiche e facilita la correzione di errori di trascrizione nei casi di interpretazione dubbia.

¹⁵ Il fatto che sia possibile inserire delle varianti non autorizza ancora a considerare il leggio-scrittoio di FAD come uno strumento per costruire apparati critici in vista della pubblicazione in rete di edizioni scientifiche in formato elettronico, anche se quanto realizzato fino a questo momento avvicina certamente quell'ambizioso traguardo. Un modulo di apparato completo ed opportunamente verificato su esempi di tradizioni testuali concrete, infatti, renderebbe integrate in un unico ambiente tecnologico sul Web sia la consultazione di archivi digitali, sia la loro trascrizione, sia, infine, la loro edizione elettronica da parte di uno o più utenti contemporaneamente. In tal senso FAD verrebbe ad assumere la fisionomia di una piattaforma per l'edizione elettronica cooperativa.

Come è stato anticipato, i materiali che sono stati oggetto di sperimentazione si riferiscono a categorie molto differenziate dal momento che compaiono sia documenti archivistici, sia libri, sia reperti archeologici e museali. Qui di seguito si forniscono informazioni schematiche relative ai dati che sono stati scelti e che, globalmente, hanno prodotto un totale di circa 300000 immagini.

4.1 *Fondazione Primo Conti*

I materiali dei Fondi Emilio Notte e Lucio Venna Landsmann sono stati scelti in base a criteri di assoluta rilevanza scientifica e culturale ed hanno generato una base dati di immagini di oltre 7000 unità. Per quanto riguarda la sezione 'Manoscritti', la partecipazione al progetto ha offerto l'opportunità di ultimare la fase di ordinamento, inventariazione e schedatura informatizzata, nonché la scansione ragionata dei documenti, che ha generato oltre 600 *record* e circa 3000 immagini. La sezione 'Fototeca', infine, ha prodotto oltre 700 *record* e circa 3000 immagini.

Non si deve dimenticare che un fatto molto positivo del progetto è consistito nel realizzare la conversione automatica di tutto il database dell'Archivio della Fondazione (circa 17000 *record*) verso il sistema FAD, facilitando così l'indispensabile aggiornamento tecnologico degli strumenti informatici ormai divenuti obsoleti.

Al fine di mettere a disposizione materiali utili per verificare le funzionalità delle componenti di fruizione specialistica del leggito/scrittoio digitale e di contribuire alla implementazione dei codici di marcatura secondo lo standard TEI, sono state eseguite le trascrizioni di alcuni documenti tratti dalla sezione 'Manoscritti'.

4.2 *Fondazione Rosselli*

L'Archivio Rosselli si presenta come un complesso documentario di straordinario valore e interesse. In esso sono raccolti manoscritti, dattiloscritti, pubblicazioni, carteggi e raccolte di giornali appartenenti a vari componenti della famiglia Rosselli e a numerosi altri personaggi e famiglie attive economicamente, culturalmente e politicamente nella storia italiana dell'Ottocento e del Novecento, entrate a vario titolo in relazione con i Rosselli (famiglie come i Nathan, Pincherle, Moravia, Lombroso, Orvieto; personaggi di grandissima rilevanza in ambito storico-politico e culturale quali Giuseppe Mazzini, Gaetano Salvemini, Maurizio Quadrio, Ferruccio Parri, Natalia Ginzburg, Sibilla Aleramo). Vi sono inoltre carte familiari, amministrative, patrimoniali contabili e le stesure dei racconti e dei testi teatrali di Amelia Rosselli.

L'archivio comprende, inoltre, una collezione di circa 150 autografi dei

maggiori esponenti storici, politici e intellettuali dei due secoli passati, che Nello Rosselli ricevette in eredità dalla famiglia materna (per esempio, Napoleone Bonaparte, Camillo Cavour, D'Annunzio, Gioberti, Giolitti, Mazzini, Pellico, Pisacane, Rossini, Ricasoli, Sella, Zola).

La prima fase di intervento sul materiale archivistico è stata realizzata utilizzando il *software* «Guarini Archivi» della Regione Piemonte per la schedatura archivistica, poi, è stata riversata in FAD al fine di procedere molto più agevolmente alla scansione e trascrizione di una parte significativa della documentazione scelta. In totale sono stati digitalizzati 2585 documenti archivistici, i quali hanno prodotto 2721 immagini. Per la fase di avvio e di sperimentazione sono state utilizzate unicamente le carte di Amelia Rosselli poiché presentavano minori rimaneggiamenti e minore discordanza, per consistenza e contenuti, rispetto agli elenchi che la Fondazione ha ricevuto in eredità insieme all'archivio.

Si è contestualmente avviato il lavoro di analisi storica finalizzata alla realizzazione della scheda-fondo (ISAD) e delle schede-soggetto-produttore (ISAAR). Il riversamento del materiale archivistico sul sistema FAD permette di rendere largamente fruibili agli studiosi documenti molto delicati, oltre che preziosi, consentendo la conservazione del materiale originale (una volta terminato completamente il riordino archivistico) nel modo ottimale e senza danneggiamenti.

Al termine del lavoro di catalogazione e digitalizzazione, FAD permetterà una esplorazione estesa al vasto orizzonte della cultura politica italiana, in particolare risorgimentale e liberal-socialista. Il risultato, che già al termine della fase di sperimentazione può dirsi importante e significativo, sarà rappresentato da un archivio ipertestuale che consentirà di mettere in relazione i materiali dell'archivio Rosselli con quelli esistenti presso altre biblioteche, archivi e musei sia torinesi, sia di altre città, idealmente collegati all'azione e al pensiero di Nello e del fratello Carlo ed all'opera artistica ed intellettuale di Amelia Rosselli.

La trascrizione dei documenti, attualmente realizzata solo su un gruppo ristretto, permetterà alla Fondazione Rosselli di offrire agli studiosi l'opportunità di conoscere nel dettaglio tutto il patrimonio archivistico di cui dispone, e di favorire lavori specialistici di edizione di materiali attualmente poco noti.

4.3 Gabinetto G. P. Vieusseux

La letteratura di viaggio costituisce per il Gabinetto Vieusseux un fondamentale filone di studio per la testimonianza che offre su molteplici aspetti, dai grandi fatti storici e culturali ai minuti particolari della vita quotidiana. Fin dalla sua fondazione, nel 1819, lo stesso Vieusseux de-

dicò particolare attenzione a far giungere tempestivamente nel suo Gabinetto di Lettura le principali opere di tale genere, ritenute indispensabile strumento di conoscenza e di aggiornamento. Il Fondo 'Fiammetta Olschki', acquistato dalla Regione Toscana e depositato presso l'Istituto nel 1981, ha rappresentato un importante incremento di tale patrimonio. Si tratta di una raccolta, prevalentemente sette-ottocentesca, comprendente 741 opere di viaggio in tutta Europa dal XVI al XX secolo, in gran parte illustrate, rappresentativa delle attenzioni e degli interessi europei. In seguito al suo arrivo al Gabinetto Vieusseux, la raccolta è stata microfilmata e ne è stato realizzato un catalogo a stampa costituito, oltre che dalle schede bibliografiche e da un ricco corredo di commento, da un apparato di indici dei nomi degli autori, curatori, artisti, traduttori, indice geografico e indice iconografico.

Grazie alla disponibilità di questi strumenti si è ritenuto di iniziare da questo fondo la digitalizzazione delle opere di viaggio presenti al Gabinetto Vieusseux, ponendo al centro del progetto la messa a punto di un prototipo di archivio iconografico delle illustrazioni di viaggio dell'Ottocento, in vista di un più ampio progetto del Centro Romantico per un «Archivio iconografico in rete sulla civiltà europea del XIX secolo». L'ingente patrimonio di immagini a stampa contenute nei volumi conservati presso l'Istituto, per qualità e quantità, può infatti rappresentare, se opportunamente contestualizzato, una fonte informativa e un documento storico di primaria importanza sulla civiltà del XIX secolo.

Attraverso l'analisi dei principali modelli di scheda in uso e il confronto con gli standard internazionali e quelli proposti dagli istituti ministeriali centrali italiani, nell'ambito del progetto FAD è stato elaborato e messo a punto un tracciato di scheda catalografica specifico e innovativo per i materiali in oggetto, che permettesse di conservare e valorizzare la ricchezza dei dati contenuti nelle illustrazioni, senza rinunciare a una potenziale condivisione delle informazioni con altri archivi. Particolare attenzione è stata riservata alla soggettazione delle illustrazioni, per la quale, nella scheda, sono stati previsti diversi sottocampi al fine di permettere una ricerca agevole a più livelli sulla molteplicità di aspetti offerta dalla letteratura di viaggio, sia attraverso liste indicizzate di soggetti, categorie e vocabolari controllati, sia attraverso gli elementi di una descrizione libera del contenuto iconografico. L'interrogazione del *database* permette all'utente finale una ricerca incrociata su tutti i campi della scheda tale da richiamare una specifica illustrazione, così come una serie di illustrazioni tra loro collegate in base a dati di contenuto storico, geografico, artistico e formale. Il tracciato è stato concepito in modo da poter consentire in futuro la messa in relazione delle

illustrazioni con le parti di testo corrispondenti, mentre già adesso le illustrazioni compaiono nel contesto delle immagini digitali dei volumi.

Sono state al momento digitalizzate 400 opere per un totale di 270.000 pagine, mentre la sperimentazione dell'archivio iconografico è stata condotta su un campione di 350 illustrazioni. Il lavoro potrà in seguito estendersi all'intero fondo e all'intero patrimonio di opere di viaggio conservate presso l'Istituto.

4.4 Istituto papirologico G. Vitelli

In seguito all'ampliamento del materiale d'archivio messo a disposizione del progetto FAD da parte dell'Istituto Vitelli, rispetto all'iniziale proposta che prevedeva non più di 1000 unità, si è portata a termine la catalogazione e la scansione digitale di quasi 7000 documenti. Essi consistono in: documenti relativi a operazioni di scavo e all'acquisto di papiri (n. 54); documenti amministrativi (n. 81); archivio Girolamo Vitelli (n. 24). A questi sono stati aggiunti sia un campione di reperti (100 papiri; 129 stoffe; 700 oggetti archeologici), sia riproduzioni fotografiche (90 foto in b/n; 28 negativi; 2697 diapositive in formato 6x6; 3566 diapositive 24x36).

La possibilità di utilizzare FAD come sistema di lettura e trascrizione di testi ha consentito di eseguire operazioni di editoria elettronica che risultano particolarmente importanti in un settore di elevata specializzazione, quale la papirologia greca. In questo ambito, le funzioni di leggio/scrittoio digitale si sono rivelate di estremo interesse anche per il fatto che aprono prospettive concrete verso la realizzazione di un sistema completo per l'edizione di testi papiracei in rete con il contributo cooperativo di più studiosi contemporaneamente.

5 Esempi di interrogazioni e ricerche sui dati FAD

Per quanto riguarda le operazioni di recupero dell'informazione sui cataloghi informatizzati e le corrispondenti immagini, trascrizioni, annotazioni ed altri eventuali dati che siano stati immessi durante l'esecuzione del progetto, saranno ora mostrati alcuni esempi per la maggior parti relativi all'archivio della Fondazione Conti¹⁶.

¹⁶ La Fondazione Conti ha fatto uso del *software* di catalogazione TINLIB, per i libri, e della sua variante TINCART, per i carteggi, sistema del quale il personale archivistico

Le ricerche possono essere condotte intervenendo sia sui valori dei campi della scheda catalografica sia sui testi delle trascrizioni.

5.1 Operazioni di ricerca sul catalogo

Relativamente alla prima modalità, si possono conoscere tutti i valori già presenti in un determinato campo, ricorrendo ad un bottone (LISTA) e si può quindi impostare una ricerca immettendo uno di quei valori (per esempio 'Italiana' nel campo 'Lingua'). FAD mostrerà tutti i documenti che corrispondono alla selezione effettuata. Sui valori dei campi della scheda catalografica è consentito di operare anche con ricerche incrociate grazie agli operatori booleani AND o OR.

I documenti trovati vengono elencati in base al numero di catalogazione: di ciascuno di essi vengono forniti i valori relativi ai campi principali della scheda catalografica, ma è tuttavia possibile consultare la scheda nella sua interezza, che, come detto sopra, può essere automaticamente generata in formato PDF e stampata per operazioni di verifica. Se invece si desidera accedere alle immagini digitali, la schermata dei risultati presenta un bottone (IMM), che consente il rinvio a tutte le icone delle immagini digitali relative al documento scelto. Cliccando su una delle icone, FAD apre una schermata che presenta, nell'area di sinistra, l'immagine digitale selezionata e, nell'area di destra, la corrispondente trascrizione e una serie di *box* disponibili per annotazioni di vario tipo (si veda la fig. 8).

Nell'area dedicata alla trascrizione, il testo può venire digitato come in un qualsiasi programma di videoscrittura. Il sistema opera automaticamente due marcature del testo digitato: applica un numero progressivo a ciascuna parola, in modo da permettere il collegamento con la corrispondente zona-parola nell'immagine digitale¹⁷; attribuisce i marcatori appartenenti allo schema di codifica XML TEI. In particolare, applica il tag *W* (*Word*) a

era esperto ed usava da tempo. Come si è detto sopra, gli archivi TINLIB e TINCART sono stati esportati e automaticamente convertiti in archivi FAD, consentendo di accedere direttamente al processo di digitalizzazione grazie ad un'interfaccia semplice ed intuitiva.

¹⁷ Si tenga presente che l'immagine digitale non è costituita da parole, ma da zone (con precisi confini in termini di *pixel*) il cui contenuto comprende parole. Per questo motivo, relativamente alle immagini, si deve parlare di 'zone-parola' e non di 'parole'.

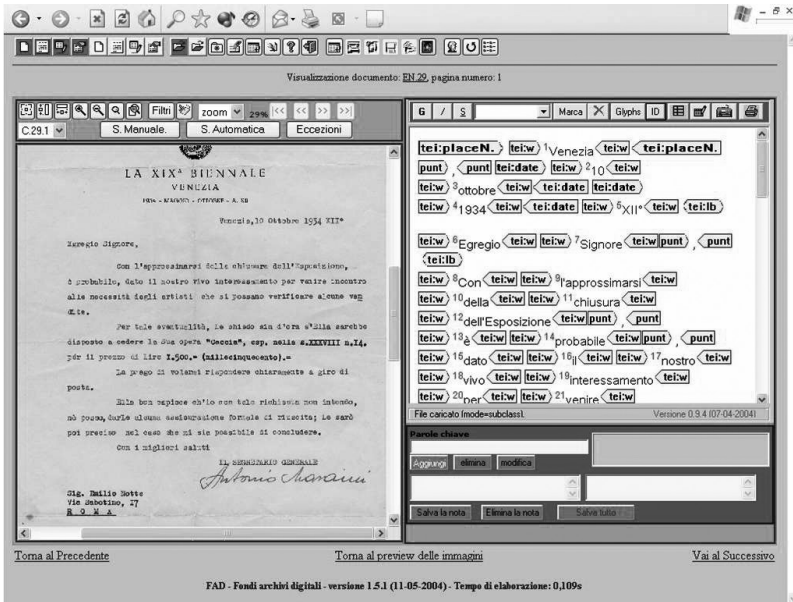


Fig. 8. L'immagine digitale e la trascrizione con le parole marcate (tag TEI) e numerate

ciascuna parola, il tag LB (Line Break) a ciascun fine linea, e il tag PUNT (Punctuation) a ciascun segno di punteggiatura¹⁸. Tale marcatura ha un ruolo essenziale, in quanto permette la visualizzazione e il trattamento del testo della trascrizione in Internet. Nella figura 8 è possibile notare come ogni parola della trascrizione sia 'circondata' dai tag W (di apertura, a sinistra, e di chiusura, a destra della parola) e sia preceduta da un numero progressivo posto in apice.

Cliccando su una zona-parola del testo entro l'immagine, la corrispondente trascrizione viene identificata nell'area di destra; egualmente, cliccando due volte su una parola nella trascrizione, la corrispondente zona-parola nell'immagine viene evidenziata. Questo collegamento biunivoco tra zona-parola nell'immagine digitale e parola nella trascrizione avviene tramite una segmentazione dell'immagine in una serie di box, ciascuno dei quali comprende una zona-parola. A ogni box è assegnato un

¹⁸ Il tag PUNT non appartiene allo schema TEI, ma è specifico della marcatura del testo operata in FAD.

numero progressivo, che permette il collegamento con la parola che, nella trascrizione, è identificata con il medesimo numero. La segmentazione può essere prodotta manualmente, oppure semi-automaticamente ricorrendo a uno specifico programma di analisi grafica che, nell'immagine, distingue lo sfondo dalle zone occupate da scrittura e, entro queste ultime, individui le eventuali colonne, le righe, e le zone-parola. Questa funzionalità può essere applicata con buoni risultati solo nel caso si abbia a che fare con documenti che presentino andamento regolare nella distribuzione delle righe di testo, e nei quali la distanza tra le parole entro le singole righe sia abbastanza ampia da permettere al programma di distinguerle correttamente¹⁹. La figura 9 riporta il risultato della segmentazione del testo presente nell'immagine di un documento della Fondazione Conti.

Oltre ai marcatori automaticamente applicati dal sistema, l'utente può aggiungerne altri a propria scelta tra quelli resi disponibili in un menù a tendina²⁰. Ad esempio, nella figura 8 è possibile vedere come alle parole «10 ottobre 1934» sia stato applicato il tag DATE (data), e alla parola «Venezia» il tag PLACEN (nome proprio di luogo).

Inoltre, entro il tag W, è possibile definire il valore degli attributi LEMMA e PoS (*Part of Speech*); la compilazione di questi valori può essere eseguita manualmente, oppure semi-automaticamente, integrando in FAD:

- un analizzatore morfologico che, analizzando i testi delle trascrizioni, riconosca gli elementi flessionali delle forme linguistiche e ad esse attribuisca il lemma dal quale derivano;
- un disambiguatore morfo-sintattico (un *PoS Tagger* e un *Parser* sintattico)²¹ che permetta di identificare quale, tra i molteplici possibili valori morfologici, sia quello corretto nel contesto sintattico in analisi.

¹⁹ Informazioni più dettagliate a questo proposito si trovano in A. BOZZI, *Testo e immagine del testo: un sistema computerizzato per le fonti manoscritte*, in *Scritti linguistici e filologici in onore di Tristano Bolelli*, a cura di R. Ajello-S. Sani, Pisa, Pacini 1995, pp. 139-172.

²⁰ I marcatori che si è deciso di rendere disponibili sono quelli che l'esperienza di marcatura manuale dei testi ha mostrato essere i più frequentemente usati. Tuttavia, la lista di questi tag può essere modificata in base a specifiche esigenze di marcatura sollevate dalle diverse campagne di catalogazione.

²¹ Un *PoS Tagger* è un annotatore grammaticale che assegna ad ogni parola di un testo la corretta parte del discorso, risolvendo le ambiguità dovute all'omografia. Un *parser* sintattico è un analizzatore che individua la struttura sintattica del testo in esame.

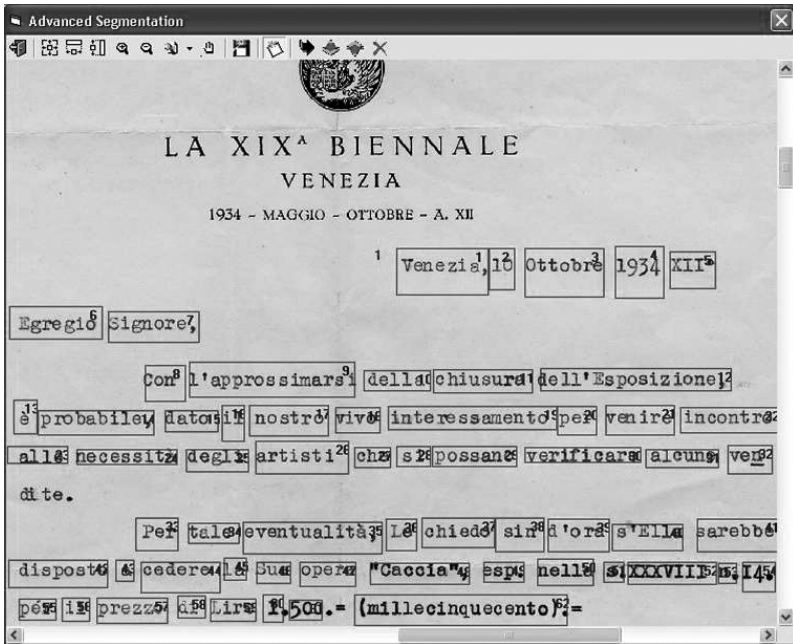


Fig. 9. La segmentazione e numerazione delle zone-parola

Nella parte bassa dell'area dedicata alla trascrizione sono presenti alcune sezioni ove riportare parole-chiave, o annotazioni relative al documento in oggetto.

5.2 Operazioni di ricerca sul testo delle trascrizioni

FAD consente di compiere ricerche sui testi delle trascrizioni: queste ricerche operano sulle forme di parola, sui lemmi, o sui marcatori. Il sistema, infatti, genera automaticamente la lista delle forme (formario), quella dei lemmi (lemmario), e quella dei *tag*: entro queste liste, l'utente può scegliere una voce (una forma, un lemma, un *tag*) e chiedere al sistema di produrre l'elenco di tutti i documenti che presentano, nella trascrizione, la forma linguistica scelta, una forma ricondotta (o riconducibile, nel caso di lemmatizzazione solo morfologica) al lemma ricercato, oppure una o più parole contenute entro il *tag* selezionato.

La figura 10 riporta i risultati della ricerca sulla forma «ottobre»: essa ricorre in due documenti (EN.29 e LVL.137), che sono, rispettivamente, una lettera e una cartolina postale. Alla destra della descrizione di ciascun documento, un bottone (LINK) permette di visualizzare l'immagine della pagina entro cui la forma ricercata appare e la corrispondente trascrizione.

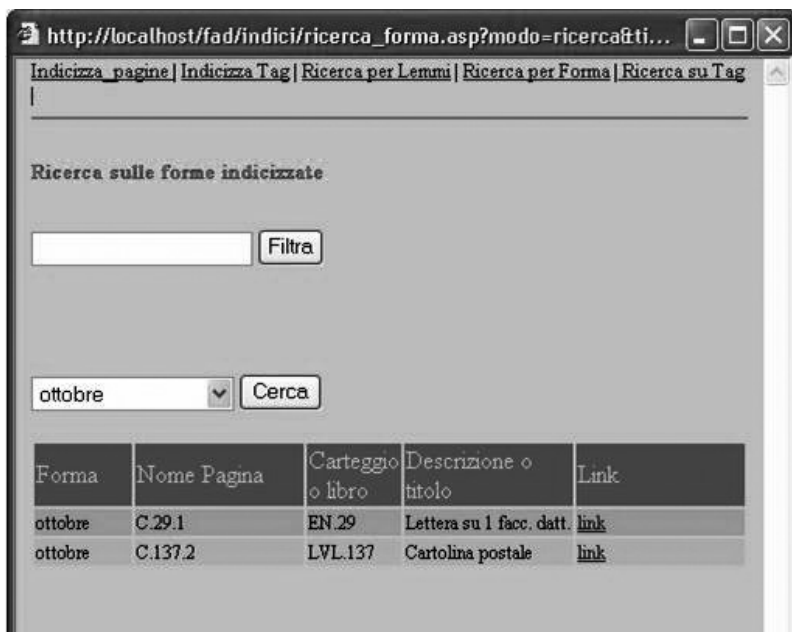


Fig. 10. Operazione di ricerca di una forma nel testo delle trascrizioni

Dal momento che FAD è in grado di gestire la catalogazione di un grande numero di documenti (fino a circa un milione nella corrente versione del sistema), sono state rese disponibili due funzioni ('Scorciatoie' e 'Segnalibri') che facilitano la navigazione. La funzione 'Scorciatoie' permette di creare fino a 5 collegamenti diretti tra pagine di documenti diversi, o pagine diverse del medesimo documento. Chi cataloga, o chi consulta, può liberamente attribuire un nome a ciascun collegamento, definendo il motivo per cui esso è stato inserito. Ad esempio, nel caso in cui due diverse lettere dello stesso autore presentino la medesima frase, o facciano riferimento a un medesimo evento, oppure si completino l'una con l'altra, potrebbe essere utile definire un collegamento tra esse, in modo tale da poterle consultare semplicemente agendo su un bottone di rinvio.

Nel caso in cui si desideri accorpare più documenti sotto una medesima rubrica, è possibile ricorrere alla funzione 'Segnalibri'. Un segnalibro ha la funzione di rinviare a una specifica pagina di un documento. Ciascun segnalibro è costituito dal proprio nome, dal numero del documento in oggetto e dal numero della pagina in cui esso è inserito: una volta forniti questi dati, il segnalibro viene aggiunto, in forma di rinvio alla pagina definita, entro una finestra che raccoglie tutti i segnalibri in una struttura ad albero. Questa struttura è costituita da una radice

(il cui nome corrisponde a quello dell'utente FAD che sta operando) e da una serie di cartelle, ciascuna delle quali può contenere uno, o più segnalibri, oppure altre sottocartelle. Cliccando su un segnalibro, FAD rimanda all'immagine della pagina in cui il segnalibro è inserito e alla corrispondente trascrizione. La figura 11 riporta una struttura ad albero esemplificativa: in essa sono visibili la radice (ADMIN), e due cartelle (FUTURISMO e RIVISTE), ciascuna delle quali contiene due segnalibri.

Nonostante che le funzionalità di FAD vengano appieno sfruttate in presenza di trascrizioni, ovvero nel caso in cui si abbia a che fare con archivi di beni librari, esso si è rivelato uno strumento utile anche alla catalogazione di immagini digitali di beni non librari (quadri, stampe, oggetti museali, fotografie). La figura 12, ad esempio, mostra l'immagine digitale di un pettine a doppia dentellatura, proveniente dal sito archeologico dell'antica città di Antinoe situata lungo il corso del fiume Nilo e dove l'Istituto Papirologico fiorentino ha condotto numerose ed importanti campagne di scavo²².



Fig. 11. Esempio di utilizzo dell'elemento di rinvio 'Segnalibri'

²² Il reperto è stato rinvenuto nel corso degli scavi effettuati negli anni compresi tra il 1965 e il 1968. Esso è conservato presso il museo dell'Istituto, dalla cui catalogazione digitale la figura è tratta.

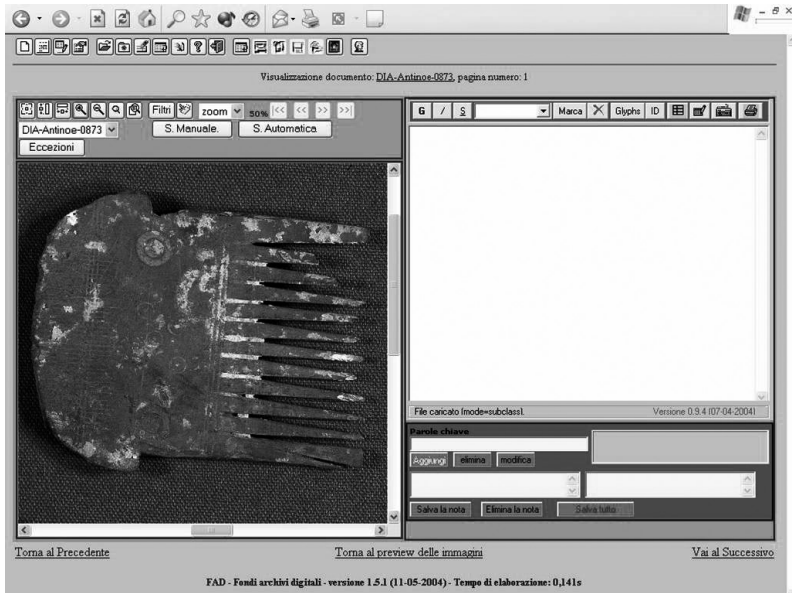


Fig. 12. Esempio di applicazione di FAD ad oggetti museali

6 Conclusioni

I sistemi che consentono la produzione di archivi digitali di grandi dimensioni sono ormai diffusi e, nonostante sia ancora dibattuto il problema della conservazione del supporto fisico sul quale le informazioni sono depositate, possiamo affermare che essi si sono rivelati affidabili e vantaggiosi, soprattutto per quanto concerne la indiscutibile semplificazione delle modalità di accesso e fruizione da parte di un'utenza senza particolari esigenze specialistiche. FAD, pur rientrando in questa categoria di prodotti, ha tuttavia dimostrato che si possono combinare, all'interno di uno stesso sistema, alcuni moduli capaci di offrire un utilizzo molto specializzato delle informazioni contenute nei documenti digitali, siano essi di tipo archivistico, bibliotecario o museale. L'architettura delle componenti tecnologiche e, soprattutto, la compresenza, in una struttura modulare, di programmi capaci di svolgere operazioni caratterizzate da forte valore aggiunto (come, per esempio, la trascrizione, l'indicizzazione e la corrispondenza fra testo ed immagine), spostano FAD verso i nuovi sistemi orientati più alla marcatura e all'estrazione del contenuto interno ai dati che alla semplice archiviazione e interrogazione dei dati stessi. In questo senso, esso considera gli istituti culturali

non solo come conservatori della memoria storica presente nei propri archivi, biblioteche e musei, ma piuttosto come produttori di contenuti che, opportunamente utilizzati, rendono assai più incisivo il ruolo culturale e formativo che essi sono chiamati a ricoprire nella società.

ANDREA BOZZI, MARCO PASSAROTTI

Tecnologie Grid e calcolo autonomico

Vedea Nembrot a pie' del gran lavoro
quasi smarrito e riguardar le genti,
che'n Sennaar con lui superbi fuoro.

Purgatorio XII, 34-36.

Le tecnologie Grid, promettendo la completa virtualizzazione di tutte le risorse informatiche disponibili, rappresentano il tentativo di risolvere gli stringenti problemi posti dalla gestione e dall'elaborazione di sempre più grandi quantità di dati. Gli sviluppi tecnologici cui abbiamo assistito sino ad oggi, culminati nell'integrazione tra Grid e Web, non sono stati sufficienti ad indirizzare e risolvere l'intricata rete di problemi dietro cui ancora si nasconde la piena potenza di una infrastruttura di Grid Computing. Considerando la complessità della Grid e la natura emergente del suo comportamento, le promesse maggiori sembrano pervenire dal paradigma del Calcolo Autonomico, supportato dagli strumenti dell'Intelligenza Artificiale.

Il problema

Caratteristica distintiva di questo inizio millennio è l'impressionante crescita della massa di dati sepolti nel silicio dei computers. Sia che la nostra attenzione sia rivolta alle indagini scientifiche, sia che il nostro interesse vada alle discipline umanistiche, dai labirintici 'omi' della Biologia Molecolare al monumentale processo di digitalizzazione del patrimonio culturale, siamo sommersi da un diluvio di bytes.

Il flusso annuale di informazioni attraverso i canali elettronici¹ che permeano il pianeta si aggira intorno ai 18 Esabytes², un terzo dei quali

¹ Radio, televisione, Internet e, soprattutto, telefono.

² <<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>>.

viene memorizzato con l'ambizione titanica di una successiva rilettura ed analisi. La difficoltà di tale rivisitazione è evidente, a partire dalla stessa unità di misura utilizzata per quantificare i dati annualmente accumulati. Un Esabyte corrisponde ad un miliardo di miliardi di bytes, suggestivamente pari a 1.470 miliardi di volte i circa 680.000 caratteri che compongono la *Divina Commedia*.

Siamo in presenza di una straordinaria abbondanza di informazioni, la cui trasformazione in conoscenza, intesa nella sua accezione più alta di bene pubblico globale, è tutt'altro che scontata. Le informazioni si nascondono più di quanto non si mostrino nel Web; pericolosamente incerte quando prive di adeguata certificazione, sono inseguite dai motori di ricerca dei grandi portali statunitensi³, sensibilmente esposte per questo al rischio di un filtro culturale. È difficile sottrarci alle grottesche suggestioni della fantasia di Elias Canetti; le tenebre in cui vive il protagonista di *Autodafè*, rischiarate solo dall'incendio della Grande Biblioteca, divengono allarmante metafora del destino del ricercatore nell'epoca dell'informazione, ma non della cultura, globale.

La conoscenza diventa bene pubblico globale col verificarsi delle tre condizioni di non rivalità, non esclusione e globalità. Ferma restando l'ineludibilità di un problema economico di base, un aiuto importante su questa triplice ed erta via può venire dalle nuove tecnologie informatiche, sia in termini di possibilità di condivisione di informazioni, che di potenza di calcolo necessaria alla loro elaborazione. Non basta saper gestire gli Esabytes, dobbiamo raggiungere i Teraflops⁴. Molte sono a questo riguardo le aspettative riposte nell'High Performance Computing, che trova in esse la sua principale ragione di essere. Per tutti l'esempio di Big Blue, il supercomputer che IBM sta realizzando appositamente per la Bioinformatica, capace di prestazioni straordinarie, con una velocità di calcolo di 300 Teraflops.

Le difficoltà non si esauriscono qui. La permeabilità tipica dei sistemi di distribuzione dell'informazione digitalizzata rende acuto il bisogno di approcci interdisciplinari, unici capaci di affrontare i dati nella loro ricchezza semantica. Una volta di più è importante il contributo che può

³ Il 60% dei 500 milioni di accessi giornalieri al Web avviene attraverso portali statunitensi. I nomi sono quelli celeberrimi di AOL, MSN, Yahoo!

⁴ Un Teraflop misura la straordinaria velocità di calcolo di mille miliardi di operazioni, con numeri non interi, al secondo.

venire dall'Informatica tramite la realizzazione dell'e-Science, supportando la funzionalità delle Organizzazioni Virtuali [1], con la rete a fare da elemento unificatore di distinte situazioni di ricerca e di applicazioni tecnologiche.

La visione

Tutti i problemi sopra elencati, necessità di grandi potenze di calcolo, capacità di gestire enormi moli di dati, abilità nel supportare le Organizzazioni Virtuali, possono trovare adeguata risposta nella visione della Grid [2, 3] un immane computer virtuale, al quale è possibile accedere, in tutta libertà, senza dover disporre di particolari conoscenze informatiche. In questa visione si diverte e si estenua, sublimandosi, la natura 'virtualizzante' dell'informatica, tesa a mostrare di ogni sistema solo le funzionalità di alto livello, in maniera standard ed interoperabile, nascondendo i dettagli della sottostante implementazione.

Si possono trovare più definizioni di Grid, ciascuna attenta a catturare un particolare aspetto dell'infrastruttura a cui si riferisce o degli scopi che con essa ci si propone di raggiungere; nella maniera più generale possibile, possiamo parlare di una Grid come di «un sistema capace di coordinare risorse che non sono soggette ad un controllo centralizzato, usando protocolli ed interfacce aperte, standard e general-purpose, con lo scopo di fornire una elevata qualità di servizio» [4].

Qualsiasi definizione, luminosa nel suo tentativo di generalizzazione, si concretizza solo nell'ombra delle tecnologie che via via si rendono disponibili; ecco che l'idea di Grid assume nomi diversi che, in diversi ambiti applicativi, ne indicano una comune funzionalità; il tutto a seconda della potenza e della connettività delle singole risorse di calcolo imbrigliate nella rete.

Al diminuire delle potenze di elaborazione dei singoli nodi della Grid, passando dai grandi centri di calcolo universitari ai comuni computers da scrivania, dalle automobili in corsa sulle strade agli onnipresenti telefonini cellulari, sino ai nostri stessi vestiti, la rete si infittisce ed il Grid Computing diventa 'ubiquitous', 'pervasive', 'mobile', 'wearable' [5-9]. Ancora, le meraviglie della Biologia dei Sistemi ci autorizzano ad immaginare un computer cellulare, la cui coerenza formale sia il risultato di una miriade di componenti di per sé inaffidabili, connessi in maniera irregolare, variabile nel tempo e fondamentalmente sconosciuta, secondo i dettami dell'amorphous computing' [10].

La visione, pur con tutta la sua carica di futuro, non è nuova e ne-

anche la sua implementazione; possiamo ritrovarne le prime tracce in silico già a metà degli anni '60, nei registri del sistema MULTICS di Corbatò e Vytosky [11]. Negli ultimi anni i venti della rivoluzione microelettronica hanno gonfiato le vele dell'idea: i nostri computers sono sempre più veloci e sempre più velocemente scambiano tra loro informazioni. Rispettando sorprendentemente i dettami della legge di Moore [12], nel periodo 1986-2000 la velocità di calcolo è cresciuta di 500 volte e quella delle reti di ben 34.000 volte. Se per la legge di Moore la velocità di calcolo del computer raddoppia ogni 18 mesi, la densità di immagazzinamento dei dati e la velocità del loro trasferimento in rete non sono da meno, anzi. La prima raddoppia ogni 12 mesi, la seconda ogni 9. Così, nel periodo 2001-2010, la velocità di calcolo è attesa crescere di 60 volte, e di 4.000 volte quella delle reti [13]. In un simile contesto tecnologico, la profezia di Leonard Kleinrock, che già nel 1969 intravedeva una macchina disgregata attraverso la rete, assume una non più remota concretezza.

L'avanzamento tecnologico si manifesta nelle due anime della Grid, quella istituzionale ed imperativa delle connessioni dedicate e quella anarcoide e volontaristica, dispersa sulla ragnatela di Internet. La Tera-Grid della National Science Foundation, al costo di circa 100 milioni di dollari, distribuisce con una rete dedicata da 10-30 gigabit/s la potenza di 9 grandi centri di calcolo statunitensi, per una velocità di 40 Teraflops ed una memoria di 2 Petabytes⁵.

Il progetto SETI@home cerca la vita intelligente nello spazio estremo⁶, affidando i propri calcoli a quanti, connessi ad Internet, siano disponibili a prestare qualche ciclo di lavoro del proprio processore. Sfruttando l'architettura BOINC⁷, un server centrale distribuisce ad una rete di uguali (peer-to-peer) un pacchetto contenente pochi dati insieme al programma per la loro analisi, aspettando fiducioso il risultato. In una giornata circolano mediamente oltre 5 milioni di pacchetti, per una esorbitante prestazione complessiva di 220 Teraflop. Assolutamente confrontabile con il picco di 360 Teraflops promesso dal prossimo mastodontico supercomputer⁸ di IBM.

⁵ Queste prestazioni si riferiscono all'ormai lontano settembre 2004.

⁶ SETI sta per Search for Extraterrestrial Intelligence.

⁷ BOINC sta per Berkeley Open Infrastructure for Network Computing.

⁸ Si tratta di BlueGene/L, creato appositamente per risolvere i problemi della Biologia Molecolare, portando agli estremi le possibilità del calcolo distribuito.

La tecnologia in atto

Il passaggio tecnologico dalla visione all'implementazione pratica è tutt'altro che semplice ed ancora incompiuto nella sua completezza. L'azione da svolgere riguarda contemporaneamente l'architettura dell'infrastruttura di calcolo e il metodo di programmazione di quella, il tutto in funzione della specifica applicazione da sviluppare.

La strada battuta è quella del middleware [1, 3], lo strato *software* interposto tra la specificità della singole risorse e la visione generale della struttura, così come questa deve idealmente offrirsi all'utente; è una sorta di traduttore universale, capace di riportare l'armonia nella confusione babelica di programmi, memorie, processori, cavi, dati ed interfacce che costruiscono la Grid. Assolve alla sua funzione garantendo i servizi fondamentali per l'integrazione delle risorse: quelli di sicurezza, dall'autenticazione alla autorizzazione; quelli per l'esecuzione del lavoro, dall'allocazione delle risorse alla gestione del workflow; quelli di informazione, dall'accesso ai dati dinamici alle sorgenti multiple; quelli per il trasferimento dei files, dalla replica al filtro [3, 14].

La frenetica corsa alla Grid ed al middleware perfetto è coordinata dal Global Grid Forum⁹, un consorzio di università, laboratori di ricerca ed aziende, che fornisce specifiche tecniche e linee guida a quanti operano nel settore. Da queste linee guida è nato un middleware standard, il Globus Toolkit¹⁰.

La tecnologia del middleware ha conosciuto una rapida evoluzione ed una altrettanto rapida convergenza verso le tecnologie Web, sino alla coniugazione di Semantic Web¹¹ e Semantic Grid¹² nell'approccio alla gestione della conoscenza. Sull'esempio dei Servizi Web, i Servizi Grid consentono di realizzare una architettura estremamente semplice ed efficace. Poggiando su questa idea, l'ultima versione di Globus Toolkit (GT4) dispiega la propria mastodontica struttura – per questo assai poco adatta a federare le modeste risorse di calcolo di un desktop PC – su una architettura Service Oriented, implementando i dettagli tecnici previsti dal Web Service Resource Framework [15-16]. Sull'esempio di quanto già avveniva nel Web, il Servizio Grid altro non è che un componente

⁹ <<http://www.globalgridforum.org>>.

¹⁰ <<http://www.globus.org>>.

¹¹ <<http://www.semanticweb.org>>.

¹² <<http://www.semanticgrid.org>>.

software, indipendente dalla piattaforma su cui agisce, e della cui esistenza siamo informati da una apposita dichiarazione, rilasciata su un registro dal provider che lo fornisce. Seguendo le informazioni depositate sul registro, possiamo collegarci direttamente al provider per usufruire del Servizio, nel rispetto delle regole dettate dal provider in relazione ad esso. Al prezzo di una architettura non perfettamente distribuita, collasata per gravitazione qua e là sui registri dei Servizi, si fondono le opere di standardizzazione della Grid e del Web [17], mettendo per di più a disposizione della prima le più mature tecnologie del secondo.

Una tecnologia in potenza

Le pur brillanti esperienze sin qui condotte nel campo delle tecnologie del middleware non hanno risolto nella loro totalità i problemi che ancora impediscono la completa virtualizzazione delle risorse disponibili. Non dimenticando l'importanza di aspetti quali la sicurezza, i portali, la gestione dei dati e dei servizi, i nodi da sciogliere per porre in atto la visione sistemica della Grid sono quelli della scalabilità, dell'adattabilità e dell'affidabilità.

L'effettiva scalabilità della struttura costituisce un obiettivo sempre più difficile da raggiungere al crescere delle sue dimensioni; è necessario decentralizzare quanto più è possibile i servizi del middleware, eliminando i residui di centralizzazione impliciti nella filosofia del 'Service-Oriented', rinunciando ai registri ed alla presenza di nodi 'più uguali' degli altri in una rete di uguali¹³.

Problemi sempre più complessi vengono affidati ad infrastrutture di calcolo le cui risorse – integrate e virtualizzate dalla Grid – sono in prospettiva sempre più numerose, eterogenee ed interconnesse; in queste condizioni, adattabilità ed affidabilità sono obiettivi di progetto sempre più ambiziosi.

Poiché la disponibilità delle risorse fluttua in seguito al transitorio apparire delle risorse di calcolo, il sistema deve adattarsi ai cambiamenti, e deve farlo in maniera affidabile, tollerando i malfunzionamenti che non possono essere eliminati – che sono anzi la cifra distintiva della Grid –, uscendo senza danno dalla situazione di stallo da questi prodotta.

¹³ Nelle reti 'peer-to-peer' ad architettura ibrida si parla esplicitamente di nodi 'super peer'.

In definitiva il nodo da sciogliere, vera ‘croce e delizia’ del Grid Computing, riguarda la complessità dell’infrastruttura che può realizzarlo. In tale complessità stanno i potenziali vantaggi del Grid Computing e le concrete difficoltà della sua implementazione.

A far fronte a questo aspetto si innalza un’altra visione, quella biomimetica del Calcolo Autonomico [18]. La Natura ha realizzato i sistemi più complessi, ed al contempo più affidabili e robusti, di cui siamo a conoscenza; rispetto a questi le più ardite opere dell’ingegno umano rappresentano una flebile ombra. Basta confrontare il volo di una mosca con il viaggio interplanetario di un satellite artificiale, per renderci immediatamente conto della superiorità del primo in termini di complessità, affidabilità, robustezza ed efficienza.

In questa visione mutuata dalla biologia, il sistema di calcolo diviene autonomico, esattamente come il battito del nostro cuore o il fluire del nostro respiro. Il ‘self’ anglosassone ne diviene la chiave di lettura; il sistema in grado di assolvere ai compiti del Calcolo Autonomico deve essere:

- self-protecting, proteggendosi da solo da accessi non autorizzati, risolvendo così ogni problema in tema di sicurezza e privacy;
- self-optimizing, gestendo da solo le proprie risorse, nella maniera migliore possibile, anche in maniera proattiva;
- self-healing, individuando e risolvendo da solo i problemi, sia di natura *hardware* che di natura *software*, che si presentano durante il suo funzionamento;
- self-configuring, autodefinendosi ‘on the fly’, senza l’intervento di un gestore esterno al sistema.

Ogni visione resta tale senza il supporto di una tecnologia; così come il middleware può costruire la Grid, la tecnologia degli agenti *software* e taluni strumenti di Intelligenza Artificiale possono costruire un sistema di Calcolo Autonomico. Combinando le diverse tecnologie possiamo indirizzarci verso un ‘middleware intelligente’ che affronti in maniera ottimale i difficili obiettivi di scalabilità, adattabilità ed affidabilità, proiettando l’applicazione su una topologia di rete realmente distribuita, integralmente ‘peer-to-peer’ [19].

Nel tentativo di conciliare la natura olistica della Grid – vista come sistema complesso – con l’inevitabile riduzionismo degli approcci ingegneristici, la chiave di volta dell’intera infrastruttura può essere rappresentata dall’agente, caratterizzato dallo stesso ruolo affidato da Koestler all’olone [20]. Come per lo studio di qualsiasi altro sistema complesso, i tre momenti di decomposizione, astrazione ed organizzazione [21] della Grid possono essere affrontati con la tecnologia degli agenti.

In senso assolutamente generale, un agente è un'entità autonoma in grado di percepire l'ambiente in cui è immersa e su questo è in grado di agire [22]. In un sistema multi-agente un numero grande a piacere di queste entità interagiscono insieme al fine di raggiungere un obiettivo comune; proprio come l'olone di Koestler, l'agente ha le due facce di Giano, deve affermare se stesso rapportandosi agli altri, nella cornice finalistica dello scopo comune. Nel caso degli agenti *software* questo brulicante sciame di entità è costituito da programmi e dati in essi incapsulati. Utilizzando gli strumenti dell'Intelligenza Artificiale, combinando le reti neurali artificiali con la logica fuzzy [23], è possibile forzare gli agenti *software* ad apprendere e modificare i propri comportamenti, sia verso l'ambiente in cui sono immersi che verso gli altri agenti con i quali vengono ad interagire, autonomamente rispetto a qualsiasi controllo centralizzato.

La natura 'organica', flessibile e distribuita dei sistemi multi-agente [24], privi di leadership ma segnati da uno scopo globale, sembra così indicare la strada verso la soluzione tecnologica ai problemi della Grid, inglobando i Servizi negli agenti e legando indissolubilmente il suo destino a quello del Calcolo Autonomico.

Conclusione

Se la visione della Grid ha già modificato la maniera di pensare alle infrastrutture informatiche, la sua implementazione tecnologica sarà sicuramente una realtà ineludibile nell'immediato futuro, con importanti implicazioni sociali e culturali. Il suo sviluppo si accompagna ad una esplosiva carica progressiva, promettendo una efficace distribuzione delle risorse «da ciascuno secondo le sue possibilità, a ciascuno secondo i suoi bisogni», in stretta ottemperanza all'adagio marxiano. Già si sono mostrate le sue potenzialità di calcolo con applicazioni 'cooperative', volte ad utilizzare le molte e piccole risorse di calcolo presenti sul Web per progetti di utilità scientifica e sociale¹⁴, senza la necessità di passare per i grandi centri di calcolo attraverso reti dedicate. La 'Grid di tutti' passa attraverso la realizzazione di un sistema in grado di automantenere-

¹⁴ Per tutti i il progetto *folding@home* per il calcolo della struttura 3D delle proteine, ed il progetto *fightAIDS@home* per la ricerca di nuovi farmaci.

si, adattandosi dinamicamente ad esigenze sempre nuove, utilizzando risorse piccole ed inaffidabili. La difficoltà è quella formidabile di progettare e costruire un sistema complesso. Abusando dell'immagine suggestiva con cui Ficino cinque secoli fa ha immortalato l'Anima Mundi, occorre far sì che il falegname alloggi nel legno. Questo, a mio avviso, spinge la ricerca verso un procedimento biomimetico ed un approccio massivamente interdisciplinare, facendo confluire la discussione sulla Grid nel più vasto alveo dell'epistemologia della complessità.

ANDREA RIPOLI

Bibliografia

1. I. FOSTER-C. KESSELMAN-S. TUECKE, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*, in «International Journal of High Performance Computing Applications», XV, 2001, pp. 200-222
2. I. FOSTER-C. KESSELMAN, *The Grid: Blueprint for a New Computing Infrastructure*, San Francisco, Morgan Kaufmann Publishers 1998
3. I. FOSTER-C. KESSELMAN-J. NICK-S. TUECKE, *The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration*, Open Grid Service Infrastructure WG, Global Grid Forum, June 2002, <<http://www.globus.org/alliance/publications/papers/ogsa.pdf>>
4. I. FOSTER, *What is the Grid? A three Point Check List*, Grid Today, July 22, 2002. <<http://www.gridtoday.com/02/0722/100136.html>>
5. P. GUPTA-D. MOITRA, *Evolving a Pervasive IT Infrastructure: A Technology Integration Approach*, in «Pers Ubiquit Comput» VIII, 2004, pp. 31-41.
6. R. BRISCOE, *The Implications of Pervasive Computing on Network Design*, in «BT Technol J», XXII, 2004, pp. 170-189.
7. P. BULL-R. PAYNE, *Pervasive Home Environments*, in «BT Technol J», XXII, 2004, pp. 65-72
8. G. BILCHEV et al., *Traffimatics – Intelligent Co-operative Vehicle Highway Systems*, in «BT Technol J», XXII, 2004, pp. 73-83
9. J.E. BARDRAM, *Activity-based computing: support for mobility and Collaboration in Ubiquitous Computing*, in «Sers Ubiquim Comput», IX, 2005, pp. 312-322
10. H. ABELSON et al., *Amorphous Computing*, Communications of the ACM 2000, <<http://www.swiss.csail.mit.edu/projects/amorphous/papers/aim1665.pdf>>
11. F. J. CORBATÒ-V. A. VYSSOTSKY, *Introduction and Overview of the Multics System*, in «Proceedings of the Fall Joint Computer Conference», XXVII, 1965, pp. 185-196

12. M. WEISER, *The Computer for the 21st Century*, in «Scientific American», CCLVII, 1991, pp. 94-104
13. G. E. MOORE, *Cramming more Components onto Integrated Circuits*, in «Electronics», XXXVIII, 1965, pp. 114-117
14. I. FOSTER-C. KESSELMAN-G. TSUDIK-S. TUECKE, *A Security Architecture for Computational Grids*, Proceedings of the 5th ACM Conference on Computer and Communications Security 1998, San Francisco 1998, pp. 83-92
15. IBM DeveloperWorks, *Web Services Notification and Web Services Resource Framework*, <<http://www.106.ibm.com/developerworks/>>
16. K. CZAJKOSWSKI-D. FERGUSON-I. FOSTER-J. FREY-S. GRAHAM-S. TUECKE, *From Open Grid Services Infrastructure to WS-Resource Framework: Refactoring and Evolution*, version 1.0. <<http://www.globus.org/wsrf/>>
17. T. BERNEERS-LEE, J. HENDLER, O. LASSILA, *The Semantic Web*, in «Scientific American», May 2001, pp. 28-37
18. J. O. KEPHART-D. M. CHESS, *The Vision of Autonomic Computing*, in «Computer», I, 2003, pp. 41-50
19. A. ORAM, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, O' Reilly, 2001
20. A. KOESTLER, *Bricks to Babel: A Selection from 50 Years of His Writings, Chosen & With a New Commentary by the Author*, New York, Random House 1981
21. G. BOOCH, *Object-Oriented Analysis and Design with Application*, Reading, MA, Addison-Wesley 1995
22. N. R. JENNINGS, *On Agent-Based Software Engineering*, in «Artificial Intelligence», CLXXVII, 2000, pp. 277-296
23. M. BERTHOLD-D. J. HAND, *Intelligent Data Analysis*, Berlin, Springer 1999
24. M. WOOLDRIDGE, *An Introduction to Multi-Agent Systems*, New York, John Wiley & Sons 2002

La magia della scrittura: modelli neurolinguistici nel messaggio scritto

Il potere delle parole

Ci sono persone che ritengono ineguagliabile la forza espressiva delle immagini.

Ci sono, è vero, immagini che parlano da sole.

L'immagine avvolge, evoca, racconta con un'intensità spesso superiore alle parole.

Non sempre, però: anche nelle parole ci sono immagini, pensieri, emozioni.

«Una giovane posa seduta, la mano destra poggia sul polso sinistro, la sinistra stretta al bracciolo in legno della sedia che corre parallelo al piano dell'immagine, così come la parte inferiore non visibile del corpo». Sembra di vederla, vero?

Leggiamo poche parole, e ripesciamo nel nostro archivio d'immagini il volto pallido della Gioconda: il dipinto più famoso del mondo. E man mano che il testo incalza, i particolari si accostano nella nostra mente sino a completarne i dettagli: lo sguardo che volge a destra, l'assenza di sopracciglia che amplia la fronte, le guance piene e i capelli avvolti da un lieve velo. Come tante piccole pennellate, le parole compongono il dipinto regalandoci emozioni forti e dirette.

Studiosi d'ogni tempo hanno attribuito un enorme valore alle parole.

Sigmund Freud diceva: «in principio parole e magia erano una sola cosa, e perfino oggi le parole conservano molto del loro potere magico. Attraverso le parole ognuno di noi può dare a qualcun altro la massima felicità oppure portarlo alla totale disperazione».

«Se mi fossero tolte tutte le mie ricchezze e forze a eccezione di una e io potessi scegliere quest'ultima, sceglierei di conservare la possibilità di parlare, perché con quella riconquisterei tutte le altre»: così Daniel Webster, un vicepresidente degli Stati Uniti.

Di opinione ancora diversa Albert Merhabian, psicologo dell'Univer-

sità della California: secondo lui ciascuno di noi proietta le proprie idee e i propri sentimenti sugli altri

- per il 7% con le parole
- per il 38% con il modo in cui le usa
- per il 55% con le espressioni del viso e del corpo.

Percentuali che ridimensionano parecchio il ruolo della parola.

È questa la ragione per cui, a volte, nonostante l'interlocutore abbia detto cose sensate, non riusciamo a concordare con lui o a vivere le sue parole in modo positivo.

I disturbi della comprensione e la decodifica del messaggio

Nel processo di trasmissione di un messaggio si susseguono tre fasi:

1. decodifica del testo
2. analisi dei significati
3. previsione del messaggio.

Nella prima fase ci confrontiamo con le regole grammaticali e con l'eventuale diversa conoscenza del codice linguistico da parte degli interlocutori: se un tedesco e un giapponese parlano ciascuno nella propria lingua, non si capiscono. Ma anche all'interno della stessa lingua, sottocodici ed espressioni usate da persone diverse possono generare disturbi di comprensione.

Nella seconda fase oggetto di attenzione non è più la grammatica, ma la semantica, i significati delle parole: «Giovanni scagliò il bicchiere contro il muro e lo ruppe».

Siamo portati a pensare che sia il bicchiere a rompersi: in Giappone, però, dove le pareti sono di carta e i bicchieri d'argento, a rompersi sarebbe il muro.

La terza fase è la più delicata, perché qui pesano esperienze e pregiudizi. Quando riceviamo una lettera dal nostro amore, una mail dal capo, un telegramma dal Ministero delle Finanze, già prima di addentrarci nelle righe abbiamo un'aspettativa: spesso un po' diversa dall'intenzione del mittente. E questo basta a pregiudicare la corretta comprensione del messaggio.

Realtà, pensiero, linguaggio: i sistemi rappresentazionali

La realtà è punto di partenza comune: ciascuno di noi compie esperienze soggettive, percepisce e archivia nella memoria in modo diverso a seconda del proprio 'sistema rappresentazionale' dominante: è il canale sensoriale che prediligiamo per elaborare le informazioni, conoscere e rappresentare il mondo. La realtà infatti diventa esperienza, l'esperienza diventa pensiero, e il pensiero prende forma con il linguaggio grazie ai cinque sensi, filtri soggettivi. Ed è per questo che ognuno di noi ha la *propria* rappresentazione della realtà.

I sistemi rappresentazionali *visivo*, *auditivo* e *cenestesico* indicano qual è l'organo sensoriale privilegiato nel raccogliere ed elaborare le informazioni percepite con vista, udito, gusto-tatto-olfatto. Il risultato? Una scelta accurata, seppur inconsapevole, delle parole usate per codificare e descrivere l'esperienza stessa.

Esiste dunque un profondo legame fra esperienza, pensiero e linguaggio e, prima ancora delle parole, il comportamento è in grado di svelare il sistema dominante di ogni persona.

I *visivi* hanno un mondo fatto per immagini e di ogni avvenimento memorizzano soprattutto colori, dimensioni e distanze. Hanno tono e volume di voce alto, respirazione veloce, postura eretta e sguardo alto. La loro gestualità è descrittiva, le mani si muovono verso l'esterno, con palmi aperti e rivolti verso il basso.

Gli *auditivi* hanno un tono di voce armonioso e a volte monotono. Inclino spesso la testa verso la fonte del suono; la loro gestualità è a tempo con le parole, le mani spesso vicine alle orecchie.

I *cenestesici* hanno un tono e un volume di voce bassi, fanno lunghe pause e assaporano ogni singola sensazione. Spalle rilassate, sguardo che punta spesso verso il basso, respiro profondo.

Ogni sistema rappresentazionale ha il proprio 'vocabolario' interno: verbi, sostantivi, aggettivi, avverbi e locuzioni avverbiali che hanno specifico riferimento al canale d'ingresso privilegiato.

I visivi hanno bisogno di: *chiarire*, *focalizzare*, dare un'occhiata, avere una *visione* completa della situazione, *orientarsi* o *valutare a prima vista*.

Gli auditivi *ascoltano*, *raccontano*, *domandano*, sono in *sintonia*, *precisano parola per parola*, *rispondono a tono o sentono qualcosa di stonato* in un discorso.

I cenestesici *annusano*, *scuotono*, *gustano*, *assaporano*, *entrano in contatto*, *si inebriano di profumi*, *provano gioia*, *eccitazione ed emozioni*, *valutano a pelle* o *di pancia*, hanno *passioni viscerali*.

Conoscere il sistema rappresentazionale del nostro interlocutore è un modo efficace per entrare in comunicazione con lui e conquistarne la fiducia.

Calibrazione, ricalco e guida: il rapport

Siamo attratti da ciò che ci somiglia, dalle persone che sentiamo simili a noi e con cui ci è facile entrare in relazione e costruire un *rapport*, ovvero una relazione armoniosa.

A volte, un buon feeling si stabilisce in modo naturale. Altre volte, invece, può nascere attraverso un lavoro di ‘calibrazione’ e ‘ricalco’.

La ‘calibrazione’ è lo studio del destinatario: dei suoi atteggiamenti, dei suoi modelli di conoscenza e di rappresentazione del mondo. E la verifica dell’efficacia del messaggio trasmesso mediante l’analisi del feedback.

Il ‘ricalco’ è un processo di rispecchiamento: riproduci il comportamento dell’interlocutore, dimostrando così interesse e attenzione al suo punto di vista e al suo modello del mondo. Con le informazioni acquisite nella fase di calibrazione aggiusti il tiro, conformando le tue azioni al comportamento più appropriato alla relazione e all’interlocutore stesso.

Calibrazione e ricalco hanno prodotto il *rapport*, ossia una relazione fatta di fiducia, empatia, disponibilità: i presupposti essenziali per passare alla guida.

La guida è l’infrazione del ricalco, ossia il processo con cui smetti di riprodurre le scelte dell’interlocutore e cominci a condurlo verso la tua mappa mentale, cercando di veicolare le tue idee e i tuoi obiettivi.

Calibrazione +
Ricalco =

Rapport ‡ Guida

Calibrazione, ricalco, rapport e guida sono fondamentali anche nella comunicazione scritta.

La struttura dell'esperienza linguistica

«La mappa non è il territorio¹»: la rappresentazione della realtà non è la realtà. La realtà è oggettiva, mentre ogni sua rappresentazione è soggettiva.

Creare una mappa del mondo è una necessità. Per quanto parziale, la mappa è lo strumento per orientarsi: la realtà (R) si confronta con l'individuo, con i suoi vincoli neurologici e sociali. Così filtrata, si trasforma in una rappresentazione della realtà (RR), una mappa personale che guida le esperienze future. Il *linguaggio*, lo si è visto, traspone in parole la mappa del mondo: è dunque la rappresentazione della rappresentazione della realtà (RRR).

Secondo Noam Chomsky, il linguaggio è strutturato su due livelli: la 'struttura profonda' (legata alla *nostra* mappa della realtà) in cui si forma il significato soggettivo dell'esperienza, e la 'struttura superficiale', ossia le parole che pronunciamo quando comunichiamo.

La struttura profonda non compare quasi mai in modo chiaro ed esplicito nel linguaggio. Nel passaggio dalla struttura profonda a quella superficiale, infatti, si verificano trasformazioni su alcune parti dell'informazione.

Cosa accade? Ognuno di noi modifica la struttura profonda del linguaggio con tre processi di modellamento:

- cancellazione => elimina alcune parti dell'esperienza o dell'informazione
- generalizzazione => trasforma alcune parti dell'esperienza soggettiva in fatti universali
- deformazione => distorce alcune parti dell'esperienza.

Dunque rendiamo la realtà soggettiva, attraverso cancellazioni, generalizzazioni e deformazioni, e creiamo un nostro *modello* o *mappa* della realtà stessa. Lo studio di questo meccanismo è chiamato 'metamodello', o *modello di precisione linguistica*, perché tende a migliorare la qualità e la quantità delle informazioni disponibili ripercorrendo a ritroso i processi di modellamento: esso ci dà uno schema per sapere quali parti del discorso sono state cancellate, generalizzate o deformate.

¹ Il principio è di Alfred Korzybski, e si legge in *Science and Sanity*, Lakeville, Connecticut, The International Non-Aristotelian Library Publishing 1933.

A questo proposito, il metamodello si serve di un 'repertorio di domande' che ci permettono di capire meglio le affermazioni del nostro interlocutore, recuperando le informazioni della sua struttura profonda andate perdute nella struttura superficiale. Domande che ci permettono, dunque, di entrare nella sua mappa.

Per esempio, prendiamo la frase: «Tutti sostengono che questa non è una buona idea». C'è un soggetto non specifico, che evidenzia un processo di generalizzazione. Applicando il metamodello, possiamo cercare di entrare nell'esperienza originaria dell'interlocutore, attraverso la domanda: «Chi, precisamente, sostiene che questa non è una buona idea?». La risposta permetterà di ritrovare una o più persone specifiche alle quali l'individuo si riferisce. O, in caso contrario, lo indurrà a rivedere le proprie convinzioni.

Vediamo ora più in dettaglio i tre processi e alcune ulteriori classificazioni.

Cancellazioni

- cancellazioni semplici: è l'eliminazione di parte dell'informazione; non è enunciato o specificato a chi o a cosa si faccia riferimento. Es.: «Sono confuso»‡Da cosa? Su cosa? «Ho paura»‡Di cosa? Di chi?
- comparazione mancante: è cancellato il secondo termine di paragone. Es.: «Claudia è più simpatica»‡Di chi? «Ormai è troppo tardi»‡Per far cosa?
- verbi non specificati: non è specificato il modo in cui si genera un'azione. Es.: «Lei mi rifiuta»‡In che modo? Con quali comportamenti?
- mancanza di indici referenziali: è la perdita dei termini in cui l'azione si sta svolgendo o si riferisce. Es.: «La gente non legge libri»‡Chi in particolare? «Sento che succederà qualcosa»‡Che cosa? A chi?

Generalizzazioni

- predicati simmetrici e asimmetrici: sono verbi che indicano atteggiamenti e azioni che implicano simmetria. Es.: «Il mio capo litiga sempre con me»‡E tu litighi sempre con lui? «I miei colleghi non mi sorridono mai»‡E tu, sorridi a loro?
- quantificatori universali: tutti, nessuno, mai, sempre, ogni volta che... sono parole che rendono universale un'affermazione. Es.: «Sei sempre in ritardo»‡Proprio ogni volta? «Non sai mai scegliere il momento opportuno»‡Davvero in nessun caso?
- operatori modali: sono i verbi volere, dovere, potere, sapere, aver bisogno di, aver necessità di...; aggiungono al verbo una modalità che indica

volontà, possibilità, dovere ecc. Es.: «Non posso cambiare lavoro»‡Che cosa te lo impedisce? «Devo terminare il progetto oggi»‡Che cosa succede altrimenti?

- performativa persa: è una generalizzazione che si presenta vera in assoluto e non solo per chi la pronuncia. Es.: «È giusto scrivere come si parla»‡È giusto per chi? «Non si deve far tardi agli appuntamenti»‡Chi lo ha stabilito?

Deformazioni

- nominalizzazione: è il processo per cui un'azione della struttura profonda emerge in quella superficiale con un sostantivo invece di un verbo. Es.: «effettuare un pagamento» invece di pagare; «apportare modifiche» invece di modificare.

- lettura del pensiero: chi parla presume di essere a conoscenza delle reali motivazioni dell'altro. Es.: «So cosa ti rende felice»; «Ho visto che eri di fretta»; «So a cosa stai pensando».

- causa/effetto: è la correlazione arbitraria di alcuni eventi secondo un rapporto di causa/effetto. Es.: «Il suo comportamento mi mette a disagio»; «la presenza del pubblico mi spaventa»; «Sentirlo mi rende felice».

- equivalenza complessa: è la forzatura del rapporto causa/effetto al limite del paradossale. Es.: «Un buon fornitore fa sempre lo sconto ai suoi clienti»; «Se mi vuoi bene devi portarmi al mare»; «Se non mi fai mai complimenti vuol dire che non apprezzi il mio lavoro».

- presupposizioni: sono frasi che devono esser vere affinché un'altra affermazione abbia senso. Es.: «Anche oggi ho lavorato troppo»; «Se il mio capo diventa furibondo come l'ultima volta che gli ho chiesto un aumento, allora ci rinuncio».

N.B. per le deformazioni non sono qui indicate le domande che aiutano a recuperare dalla struttura profonda le informazioni perdute. Se hai voglia di fare una pausa nella lettura, è una buona occasione per un esercizio.

Internet: esigenze di chiarezza quando il rapporto è a distanza

Cominciamo dal dominio. Parliamo dell'*URLability*, il neologismo creato per indicare l'usabilità degli URL, ovvero degli indirizzi dei siti web. Un sito ben progettato ha un nome di dominio breve, coerente con i contenuti, facile da digitare e ricordare.

Autoironici i domini tutti al femminile: <<http://www.wow-wome->

nonwork.it> (organizzazione no profit specializzata nella comunicazione multimediale) e <<http://www.dols.it>> (mica bamboline, bensì esperte di nuove tecnologie). Ce ne sono altri, poi, che giocano sull'ambiguità fonetica come <<http://www.froogle.google.com>> (froo.gle, si legge frugal). Provocatori, infine, <<http://www.dammela.it>> e <<http://www.abbordescion.it>>.

Nel mare mediatico in cui siamo immersi, oltre all'informazione in sé, conta il contesto in cui l'informazione stessa è inserita, il modo in cui è presentata, commentata e integrata.

Chi scrive nel web è chiamato a:

- progettare pagine autonome e autoconsistenti per contenuti e funzionalità
- occuparsi di cose semplici ma non banali come il testo in home page, la data dell'ultimo aggiornamento e il copyright
- prevedere sezioni esplicative ed etichette coerenti per ogni sezione
- usare le mappe di navigazione per tracciare il percorso di conoscenza del lettore.

Metadata ben scritti (sono i testi nascosti nel codice html) favoriscono sia l'autore sia il lettore, il proprietario del sito e l'eventuale acquirente o navigatore: scegliere quelli giusti richiede abilità e conoscenza dei comportamenti del navigatore. In questa direzione, la familiarità con i modelli neurolinguistici è di grande aiuto.

Diamo ora uno sguardo alle home page di alcuni siti (dati rilevati nel marzo 2005):

<<http://www.repubblica.it>> - In uno dei siti più visitati, pesa un errore tanto banale quanto insidioso per il navigatore: i link sono in rosso o in nero (oltre al canonico blu) e non cambiano colore una volta 'visitati'. Inoltre c'è una sorta di battaglia metaforica all'interno del sito: sia nel logo che nella barra del browser si ribadisce La Repubblica, ma l'URL è <<http://www.repubblica.it>>. E se un utente digitasse <<http://www.larepubblica.it>>? Vedrebbe un enorme punto interrogativo.

<<http://www.webaccessibile.org>> - Nel sito italiano di riferimento per l'accessibilità l'iscrizione alla newsletter con la prima parte (tuamail) è imprecisa a livello informatico, anche se familiare; mentre la seconda (dominio.tld) è esatta, anche se un po' troppo da specialisti.

<<http://www.uil.it/ambiente>> - Bassa leggibilità: troppi colori, sfondo pesante, link bianchi distribuiti su più righe.

<<http://www.forza-italia.it>> – Il sito ufficiale di Forza Italia si caratterizza per una comunicazione leggera, quasi spensierata. Le testimonianze creano empatia e rafforzano la comunità.

<<http://www.lafabbricadelprogramma.it>> – Diverso il tono scelto dall'Unione per l'iniziativa 'La fabbrica del programma', che vuol essere uno spazio – fisico e virtuale – dove discutere i grandi temi del Paese. Interessante l'uso della metafora: 'fabbrica', un termine forse fuori moda diventa sinonimo di lavoro, cooperazione, spirito di squadra, condivisione di un obiettivo comune, concretezza.

<<http://www.buonalombardia.it>> – Un sito istituzionale che, lontano da formalismi e gerghi di settore, emana tutto il profumo, i colori e sapori della terra che rappresenta... è un tuffo con tutti i sensi nell'enogastronomia lombarda di qualità, che seduce il lettore grazie a un linguaggio evocativo e sensoriale. Qualche esempio:

Itinerari tematici => Crea il tuo itinerario, riempi di ciò che ami il cestino di Buona Lombardia: ti aspettano le strade dei vini e dei sapori, tutto il buono e il bello della Lombardia.

Gastronomia lombarda => Sapori, profumi e aromi della Lombardia, racchiusi in un'ampolla da aprire poco alla volta per gustare al meglio le sfumature di ogni prodotto.

Le formule della seduzione: essere abilmente vaghi

Migliorare la qualità e la quantità delle informazioni disponibili. Ma la strada del metamodello non è l'unica percorribile. Intendiamo 'informare' o 'sedurre'? In alcuni casi essere espliciti è indispensabile:

Il termine di scadenza per la presentazione delle domande di iscrizione è fissato al 9 settembre 2006.

In altri casi conviene sfumare:

È probabile ti venga la tentazione di lasciarti alle spalle le cose che ti hanno aiutato ad arrivare dove sei.

Generalizzazione, cancellazione e deformazione si possono percorrere a ritroso, per risalire all'esperienza sottesa e arricchire il messaggio, ma

possono anche essere sfruttate per mantenere la comunicazione in una fascinosa indeterminatezza, che stimola le scelte del lettore:

Chi sa leggere i tuoi pensieri sa sempre cosa vuoi...

È la strada del *Milton model*, che alla chiarezza del metamodello oppone la vaghezza di un linguaggio volutamente emozionale. Due diversi linguaggi: il primo comunica con l'emisfero sinistro del cervello e con la coscienza, e usa costruzioni grammaticali e sintattiche. Il secondo comunica con l'emisfero destro e con l'inconscio, e usa immagini, metafore, suggestioni evocative.

Un esempio di linguaggio abilmente vago e seducente: <<http://www.esperya.com>>.

È un sito che vende prodotti gastronomici online, e lo fa con arguzia e garbo, soprattutto grazie ai testi. Analizziamo il 'pacco del diffidente'.

Il testo ci strizza l'occhio già dal titolo: oltre a indicare la confezione di oggetti, *pacco* gergalmente significa 'fregatura'. Ma scioglie subito l'ambiguità, e cerca un rapporto di fiducia col lettore:

testo	analisi
Sono diffidente!	Ricalco dello stato d'animo del lettore, e cancellazione (diffidente verso che cosa?).
Non hai mai fatto un ordine su Esperya.	Presupposizione e quantificatore universale. Negazione che sospinge all'idea di <i>fare un ordine</i> .
Sei diffidente, è naturale.	Lettura del pensiero, e performativa persa, per di più con un aggettivo, <i>naturale</i> , che richiama la genuinità dei prodotti gastronomici offerti.
E io voglio (debbo, è il mio lavoro di bottegaio!) vincere la tua diffidenza, conquistare la tua fiducia.	Congiunzione <i>e</i> , che avvicina lettore e scrittore, avvicinarsi di operatori modali (inoltre <i>debbo</i> , che suggerisce un'atmosfera antica, tradizionale, e <i>bottegaio</i> , che richiama la familiarità del piccolo negozio sotto casa). Seguono due verbi, all'infinito, con il sapore dell'ottimismo.
Se sei disposto a darmi una chance e a mettermi alla prova, clicca qui, altrimenti... beh, non ce l'ho fatta!	Guida verso la soluzione voluta: domanda e suggerimento nascosti. Il potenziale cliente è <i>disposto</i> , dà una <i>chance</i> , mette <i>alla prova</i> : è lui che decide. Dall'altra parte c'è chi offre: <i>altrimenti</i> (alternativa che conferma la libertà del lettore), puntini di sospensione, interiezione familiare, ricalcano i toni di una conversazione vis-à-vis: autore e lettore sono dalla stessa parte: l'ammissione affabile <i>non ce l'ho fatta!</i> è quella dell'amico al bar. Proseguiamo la lettura, vero?

E-mail: croce o delizia?

Write the way you speak. Ma sì, l'e-mail è *written speech*, discorso scritto. Non staremo mica sempre lì a rileggere tutto prima d'inviare? In realtà, molta gente le e-mail le stampa, le archivia, ne fa fotocopie, le inoltra a persone che ritiene possano essere interessate. E ancora, spesso le e-mail sono l'unico e il più importante canale per intrattenere relazioni con i nostri pubblici: clienti, colleghi, fornitori ecc.

Strumento velocissimo, cui non è possibile apportare correzioni, l'e-mail richiede tempo, precisione, attenzione al destinatario, sicurezza ed efficacia.

Conviene per questo adottare alcune precauzioni psicologiche:

- Prima fra tutte: controllare la propria 'e-mail mania'. Il dialogo e il contatto personale restano insostituibili. Anche perché la e-mail sa rendere isteriche anche le persone più miti, che online arrivano a mandarsi al diavolo, magari per una parola che, a voce, avrebbe avuto tutt'altro effetto. Il fenomeno è noto col nome di *flaming*: la velocità e l'impetuosità tecnologica incoraggiano comportamenti di rottura che faccia a faccia sarebbero autocensurati.
- Da controllare anche le reazioni emotive: se nella mail box arriva una critica, invece di un complimento, cerchiamo di trarre vantaggi dall'interattività, rispondendo garbatamente e rimediando all'errore (se ci fanno davvero arrabbiare, sbolliamo prima di rispondere: a scrivere sull'onda della rabbia avremo sempre da pentirci, anche se abbiamo ragione da vendere).

Poi, alcune precauzioni comportamentali:

- non allegare file senza avvertire: sono lunghi da scaricare, provocano fastidio e, se arrivano da chissà dove, possono contenere virus;
- non fare spamming e non partecipare a catene di sant'Antonio telematiche;
- non abusare della funzione 'copia conoscenza': abbiamo tutti presente il delirio creato dalla possibilità di mandare un messaggio a più destinatari contemporaneamente: caselle intasate da messaggi che non ci riguardano affatto, ma che ci hanno mandato «così sei al corrente anche tu»; spesso con allegati pesantissimi; spesso con decine d'indirizzi visibili nella casella 'CC', in barba alla privacy; con il risultato che nessuno prende iniziative, convinto che a decidere dovesse essere qualcun altro della lista;
- non abusare della funzione 'copia nascosta': può creare caos di re-

sponsabilità; Tizio manda la mail a Caio, ma poco gli importa delle sue reazioni: voleva solo fargli sapere che l'ha mandata anche a Sempronio (che sta in 'CC'); magari senza far sapere a nessuno dei due che l'ha mandata in 'CCN' anche a un altro lettore misterioso – probabilmente più in alto di tutti nella scala decisionale – che osserva il chiacchiericcio dal buco della serratura, e sornione se la ride.

Infine, alcune precauzioni tecniche:

- Curare la qualità dei testi: brevi e veloci, sì; trascurati no. Lo stile telegrafico, l'andamento gergale, lo scempio all'ortografia, gli anglicismi dappertutto: *Ti forwardo una mail con attachment, droppala se non ti serve, ma replyami please* (t'invio un messaggio che ho ricevuto, con un allegato; se non t'interessa cestinalo, ma per favore rispondimi). Tutto questo può generare gravi malintesi. Una buona abitudine: non spedire un messaggio appena scritto, ma stamparlo, rileggerlo, ripulirlo, scriverci sopra i dubbi che può suscitare, riprenderlo dopo un po' sistemando i punti chiave, e infine inviarlo a destinazione. Questo dà concretezza ai concetti ed elimina molti problemi.

- Scrivere il titolo (subject) concentrando il messaggio in poche parole, cariche di significato. Quei pochi caratteri sono essenziali perché il destinatario apra il nostro messaggio. Forma breve, asciutta, senz'articolli, esplicita e precisa. Non 'novità', ma quale novità. Non 'emergenza', ma il motivo dell'emergenza. Non 'comunicato stampa' (s'immagina la quantità di mail intitolate 'comunicato stampa' che arrivano nelle redazioni dei giornali?), ma il titolo del comunicato. Non 'invito', ma 'invito al convegno su...'.

- Essere brevi: condensare il messaggio in una sola schermata, e scrivere la cosa più importante nelle prime righe. Nessuno ama 'scrollare' in fondo allo schermo per cercare la notizia. Se c'è bisogno di più spazio, iniziare con un indice per punti, così chi è interessato al punto 5 saprà subito dove cercare.

- Messaggi densi, ma rarefatti: densi di contenuto, rarefatti sul piano visivo. Leggere sullo schermo stanca molto più che leggere sulla carta. Paragrafi brevi, titoletti, spazi bianchi, liste puntate o numerate, per dare vivacità e leggerezza al messaggio.

- Sfruttare l'interattività, introducendo uno o due link nel testo: al sito appena aggiornato, al form da compilare per iscriversi al convegno, o a qualsiasi altra fonte sia ritenuta di valore per il lettore. Quelle parole colorate e sottolineate attirano la sua attenzione: e se le prime righe lo hanno incuriosito, è difficile che resista alla tentazione del clic.

- Stile informale, ma non troppo: via il vecchiume dalla rete («Con la presente siamo a informarLa...», «In attesa di un Suo sollecito riscontro...»), ma senza concedersi qui libertà che faccia a faccia non ci concederemmo.

- Usare bene la firma: non solo nome e cognome, ma anche indirizzo, telefono, fax, e-mail, sito Internet. Nella e-mail, la firma funziona come un biglietto da visita. Ed è utile anche con lettori ben noti e in buona confidenza: mettere lì tutti i riferimenti facilita il contatto, può generare una telefonata per chiarire un dubbio, consolida la relazione.

Anche nelle e-mail, inoltre, è molto utile cercare di conoscere il nostro destinatario e di instaurare un buon rapporto con lui. Anche in questo caso, dunque, ha senso un processo di calibrazione, ricalco e guida:

- calibrazione > accogliamo informazioni su chi scrive. Perché ci scrive, cosa vuole da noi, qual è il suo stato d'animo, quali scelte ce lo rivelano: scelte verbali (registro lessicale, sistemi rappresentazionali, verbi, sostantivi, aggettivi, avverbi); scelte paraverbali (tono, stile, fluidità, ritmo, punteggiatura...); scelte non verbali (grafica, formattazione, maiuscole...).

- ricalco > individuiamo quali parti del suo testo ci conviene ricalcare per entrare in buona relazione. Attenti anche alle scelte che *non* conviene ricalcare.

- guida > individuiamo l'obiettivo della relazione con il lettore e il messaggio che vogliamo dargli: creata la sintonia, muoviamoci con lui verso quella direzione. Con questo intento, scriviamo la risposta.

Consapevolezza è libertà di scegliere

Ogni scelta linguistica è, appunto, una scelta. Con la sua efficienza e fragilità. D'altronde i limiti della parola sono gli stessi che ne sanciscono la forza: di qui realtà, di là rappresentazioni soggettive.

In mezzo: la consapevolezza. Consapevolezza che sospinge al rapporto: *diventare* l'interlocutore, scrivere come vorremmo leggere, risponde alla felice logica *win/win*: io vinco/tu vinci. Vinciamo insieme.

Consapevolezza che ci permette di scegliere: quando abbandonarci agli automatismi di competenze acquisite, o quando accarezzare le parole una a una. In fondo, dipende da noi.

ALESSANDRO LUCCHINI

metodo CRG

Calibrazione

raccogliamo informazioni su chi ci scrive: perch   scrive, che cosa vuole da noi, qual   il suo stato d'animo, quali scelte ce lo rivelano:

- scelte verbali (registro lessicale, sistemi rappresentazionali, verbi, sostantivi, aggettivi, avverbi ...)
 - scelte paraverbali (tono, stile, fluidit  , ritmo, punteggiatura ...)
 - scelte non verbali (grafica, formattazione, maiuscoli ...);
- ====> senza fretta di giudicare!

Ricalco ====> rapport

individuiamo quali parti del suo testo ci conviene ricalcare per entrare in buona relazione;

====> attenzione anche alle scelte che non ci conviene ricalcare;

Guida

individuiamo l'obiettivo della relazione con il lettore e il messaggio che vogliamo dargli: creata la sintonia, muoviamoci con lui verso quella direzione; con questo intento, scriviamo la risposta.

Un grazie a Rosella Gaudiuso e ad Annalisa Pardini per l'aiuto offerto nella stesura di questa relazione.

Apprendimento automatico di ontologie da testi: selezione di relazioni semantiche da un *corpus* di biologia molecolare

1 Introduzione

Nel campo dell'*information retrieval*, lo studio dell'accesso ed elaborazione dell'informazione digitale, soprattutto su base testuale, una crescente attenzione è rivolta all'informazione contenuta all'interno dei documenti. In questa prospettiva l'unità di misura non è più il documento stesso, quanto il contenuto del documento, mentre l'obiettivo è l'individuazione di informazioni che potrebbero anche non essere espresse letteralmente nel documento, o magari contenute in più documenti. Questa area di studio coinvolge discipline che vanno dall'informatica, alle scienze cognitive e linguistiche, alla matematica e alla filosofia. L'utilizzo di tecniche di trattamento automatico del linguaggio (TAL o NLP)¹ e apprendimento automatico², per l'analisi semantica su larga scala, viene anche chiamato *text mining* (HEARST 1999), per porre un accento sull'obiettivo dichiarato della scoperta di nuova informazione dai testi.

La bioinformatica è uno dei domini di ricerca più attivi per il *text mining*, a causa degli enormi database di letteratura scientifica digitale, come SwissProt³ e PubMed⁴. La crescita di queste risorse è nell'ordine

m.ciamita@istc.cnr.it. Laboratorio di Ontologia Applicata (LOA, <<http://www.loa.cnr.it>>) Istituto di Scienze e Tecnologie Cognitive (ISTC-CNR), Via Nomentana 56, 00161 Roma. Massimiliano Ciamita lavora attualmente presso Yahoo! Research, Ocat, 0800, Barcellona - massi@yahoo-inc.com

¹ *Natural language processing*.

² *Machine learning*.

³ <<http://us.expasy.org/sprot/>>.

⁴ <<http://www.pubmed.gov>>.

di migliaia di nuovi documenti ogni mese. L'obiettivo principale del *text mining* in bioinformatica è la scoperta di nuova conoscenza biomedica. Un esempio concreto delle potenzialità di questo tipo di ricerca è il lavoro di Swanson e Smalheiser (1997), i quali mettendo a confronto, con metodi statistici, testi di biomedicina hanno individuato nuove importanti relazioni causali, per esempio fra magnesio ed emicrania.

1.1 *Text mining ed estrazione di informazione*

L'approccio maggiore emerso negli ultimi anni, per accedere automaticamente al contenuto dei documenti, è quello basato sull'“estrazione di informazione” (IE). Questo approccio si basa in primo luogo sulla definizione di un ‘modello concettuale’ del dominio di interesse. In altre parole un'ontologia, o meglio un'ontologia di dominio. L'ontologia di dominio specifica una serie di concetti rilevanti ai fini del processo di analisi dei contenuti. In particolare, definisce un certo numero di ‘classi’, e delle ‘relazioni’ le quali codificano interazioni rilevanti fra le classi definite. È importante notare che le ontologie usate in queste applicazioni normalmente non sono provviste di una semantica logica esplicita, cioè non sono ‘ontologie formali’ (cfr. GUARINO 1998). Una volta definita l'ontologia si sviluppano programmi, anche detti annotatori (o *taggers*), che devono analizzare i documenti e individuare i concetti e le relazioni definite dall'ontologia dove menzionati. Un'interessante esempio di estrazione di informazione da testi di biomedicina è quello proposto da Rosario e Hearst (2004) che implementano sistemi di annotazione automatica basati su un'ontologia costituita da due classi: *Disease* (malattia) e *Treatment* (cura), e relazioni come *cure* (cura), *has side effect* (ha effetto collaterale), *prevent* (previene), ecc. La figura 1 mostra schematicamente questo modello concettuale che può essere usato, congiuntamente a un *corpus*, per sviluppare un *tagger* che individui nei testi istanze di concetti e relazioni pre-definite, come nell'esempio seguente, in cui si identifica una relazione di cura tra una forma di epatite e un tipo di terapia:

(1) These results suggest that [con A-induced hepatitis]*DISEASE* was ameliorated by [pre-treatment with TJ-135]*TREATMENT* |*CURE*.

Normalmente le ontologie usate per l'estrazione di informazione sono di piccole dimensioni. In altre parole, definiscono un numero limitato di classi e relazioni. Questo è dovuto soprattutto al fatto che la costruzione di ontologie per applicazioni di questo tipo è un'impresa complicata e onerosa. Inoltre, le ontologie più diffuse definiscono esclusivamente

concetti e relazioni ‘strutturali’ fra concetti, cioè relazioni come «è un tipo di» (*is-a*, *is a kind of*), «parte di», ecc. ma non relazioni semantiche arbitrarie. La definizione di relazioni strutturali, inoltre, può essere supportata con successo da tecniche automatiche e semi-automatiche di apprendimento ben studiate. Mentre l’apprendimento di relazioni arbitrarie è un problema meno conosciuto.

In questo lavoro investighiamo il problema dell’arricchimento di un’ontologia esistente con relazioni semantiche arbitrarie fra i concetti definiti, e presentiamo un sistema per l’apprendimento automatico di tali relazioni da *corpora*. Il nostro sistema ha un’architettura molto semplice, in quanto utilizza una serie di strumenti di trattamento automatico del linguaggio robusti e molto diffusi. In particolare, il sistema utilizza l’analisi sintattica, la ricognizione di *named entities*, e analisi statistiche per la scoperta di collocazioni lessicali e l’apprendimento delle proprietà semantiche degli argomenti di predicati. Nella sezione sperimentale mostriamo come l’applicazione del sistema a un *corpus* di letteratura di biologia molecolare, il *corpus* di GENIA (OHTA *et al.* 2002) generi una lista di relazioni binarie ordinate, etichettate con concetti dell’ontologia; per esempio «Cell express-the-receptor-for Protein», o «Virus replicate-in Cell»⁵. Le relazioni scoperte dal sistema sono, in seguito, analizzate approfonditamente. Da questa analisi emerge che il sistema, oltre ad essere semplice, è molto accurato. Ciò lo rende potenzialmente utile per applicazioni di *text mining* in cui si voglia generare automaticamente una lista di relazioni fra concetti, o per il supporto alla costruzione semi-automatica di ontologie da parte di esperti. È ben noto infatti che la generazione di terminologie dal basso, cioè dai testi, e non solo elicitati da esperti, aumenta significativamente la copertura delle ontologie per applicazioni di information management.

Nella sezione seguente descriviamo in dettaglio il problema dell’apprendimento di relazioni ontologiche. Nella Sezione 3 descriviamo il nostro sistema e i dati degli esperimenti svolti. Nella Sezione 4 discutiamo la valutazione dell’output del sistema. Infine, nella Sezione 5, presentiamo le nostre conclusioni.

2 Relazioni in un’ontologia di biologia molecolare

⁵ Rispettivamente, «esprime il recettore per», fra «Cellula» e «Proteina», e «si replica in» tra «Virus» e «Cellula».

GENIA (OHTA *et al.* 2002) è un'ontologia di biologia molecolare che definisce una serie di classi riguardanti l'espressione di geni e la loro regolazione, reazioni segnalatrici di cellule, ecc. Una considerevole mole di lavoro in bioinformatica si è concentrata sul problema della ricognizione automatica di nomi di 'entità'⁶. La ricognizione di entità è una componente fondamentale dell'estrazione di informazione, e consiste nell'individuazione nei testi delle occorrenze di classi specificate nell'ontologia di riferimento. Questo compito si concretizza, nei sistemi più accurati e sofisticati, come il problema di assegnare a una sequenza di parole una sequenza di etichette definite dall'ontologia. Per esempio, l'obiettivo potrebbe essere di riconoscere, nella frase seguente, estratta dal *corpus* di GENIA, che «NS-Meg cells», «mRNA» e «EPO receptor» sono, rispettivamente, istanze delle classi «Cell_line», «RNA_family_or_group», e «Protein_molecule»:

(2) Untreated [NS-Meg cells]_{Cell_line} expressed [mRNA]_{RNA_family_or_group} for the [EPO receptor]_{Protein_molecule}.

L'estensione naturale del processo di estrazione di informazione consiste, come visto sopra nel caso di (ROSARIO-HEARST 2004), nella classificazione delle relazioni esistenti fra istanze di classi nel *corpus*. L'estrazione di eventi è un componente cruciale per la gestione di rappresentazioni semantiche più complesse, necessarie per le operazioni di analisi automatica del tipo proposto in *text mining*. Oggi la maggior parte della ricerca sull'estrazione automatica di relazioni utilizza pattern, o regole, per l'estrazione di relazioni costruite manualmente; per esempio Friedman *et al.* (2001) per l'identificazione di percorsi segnalatori molecolari⁷, e Šarić *et al.* (2004b) per ricercare informazioni sull'interazione di proteine, applicando una ontologia di dominio simile a quella descritta da Ratsch *et al.* (2003). Gli approcci basati su risorse e sistemi di regole costruiti manualmente tendono ad avere una buona precisione, ma offrono poca copertura, poiché riguardano, per forza di cose, pochi casi di relazioni. Gli approcci basati su sistemi di apprendimento statistico tendono ad avere sia buona precisione che copertura, come mostrato in (ROSARIO-HEARST 2004) dove metodi di apprendimento sequenziale

⁶ NER, *named entity recognition*.

⁷ *Signaling pathways*.

di named entities sono estesi all'estrazione di relazioni. Tuttavia anche questo approccio si basa sulla creazione manuale di un'ontologia e di un *corpus* annotato, ed è quindi limitato dalle risorse e il tempo che possono essere investiti per la creazione di questi strumenti. In questa area di ricerca sono state studiate tecniche per trovare automaticamente pattern di estrazione di un insieme di relazioni semantiche definite, per esempio da Craven e Kumlien (1999) riguardo a cinque relazioni binarie che coinvolgono proteine, o da Pustejovsky *et al.* (2002), riguardo all'estrazione di relazioni inibitive. Lo sviluppo di metodi per individuare relazioni cruciali fra entità definite dall'ontologia di dominio si inserisce perciò naturalmente, colmando una lacuna, in un contesto di strumenti e applicazioni complementari in IE. Infine, è da notare che relazioni semantiche di questo tipo sono state usate anche per la verifica della coerenza dei dati in database di biologia molecolare, per individuare errori e allineare database diversi, e come template, o principi guida, per la generazione di schemi di database (ROJAS *et al.* 2002).

Sistemi di *text mining* in biomedicina basati anche sull'estrazione di relazioni richiedono un insieme di relazioni predefinite. Tipicamente queste relazioni vengono definite in modo manuale da esperti di dominio e di ingegneria ontologica, con un lavoro complesso e oneroso in termini di tempo e risorse coinvolte. Questo processo, inoltre, tende verso la produzione di risorse caratterizzate da una buona qualità (o *precision*), ma, allo stesso tempo, da bassa copertura (*recall*). Nonché dall'ineliminabile *bias* determinato dalle conoscenze specifiche degli esperti coinvolti. L'obiettivo del nostro lavoro è di generare automaticamente tutte le relazioni che codificano interazioni informative, attestate in grandi *corpus* di testi, tra coppie di istanze di classi definite in un'ontologia di oggetti più semplice pre-esistente. Il prodotto del nostro metodo consiste nell'arricchimento dell'ontologia di classi con relazioni binarie direzionate, tra classi pre-esistenti. L'obiettivo è supportare i processi di costruzione di ontologie, suggerendo elementi da inserire, e supportare sistemi di *text mining* e estrazione di informazione. In entrambi i casi aumentando la copertura, vale a dire il volume dei fenomeni identificati.

La maggior parte degli studi sulla costruzione automatica di ontologie ha fin qui riguardato soprattutto l'acquisizione di relazioni ontologiche di tipo strutturale; per esempio, nuove istanze di categorie come veicoli, persone e luoghi, e perciò l'apprendimento di relazioni di tipo Is-a (cfr. per esempio HEARST 1992; PANTEL-RAVICHANDRAN 2004), oppure di relazioni di tipo parte-di (BERLAND-CHARNIAK 1999). Più di recente

Almuhareb e Poesio (2004) hanno investigato il problema dell'acquisizione di attributi di classi ontologiche, mentre Reinberger *et al.* (2004) hanno proposto un sistema per l'acquisizione di relazioni soggetto-verbo-oggetto, questo ultimo lavoro in particolare sarà discusso più dettagliatamente nella parte finale dell'articolo, in quanto riguarda la nostra metodologia più direttamente.

3 *Apprendimento automatico di relazioni da testi*

Il nostro metodo utilizza un *corpus* di testi e un'ontologia di dominio che definisce un insieme di classi e, se disponibili, relazioni strutturali fra queste classi. In questo lavoro utilizziamo l'ontologia di GENIA e il relativo *corpus* di abstract di letteratura biologica. Il primo trattamento dei testi riguarda la ricognizione di occorrenze di istanze di classi definite dall'ontologia, attraverso tecniche di named entity recognition, come nell'Esempio 2 sopra. Nel nostro caso questa fase è solamente simulata, poiché il *corpus* di GENIA è già annotato. In generale, questo pre-trattamento può essere eseguito attraverso tecniche standard di NER, che possono utilizzare il *corpus* per la fase di training. L'output del nostro metodo consiste in un insieme di relazioni semantiche binarie direzionate, fra coppie di classi di GENIA; per esempio, una relazione plausibile potrebbe essere «Virus infect Cell».

3.1 *Corpus e ontologia*

L'ontologia di GENIA è stata costruita per modellare eventi biologici a livello cellulare, e per essere usata come base per sistemi automatici di estrazione di informazione da testi di letteratura biomedica. L'ontologia consiste di 46 concetti nominali organizzati in una struttura tassonomica generica, cioè le relazioni strutturali (Is-a, parte di, ecc.) non sono chiaramente specificate; vedi fig. 2. Questa ontologia è stata utilizzata per annotare un *corpus* con etichette semantiche, relative alle istanze di classi definite nell'ontologia. Noi qui utilizziamo il *corpus* GENIA, versione G3.02, il quale consiste di 2.000 abstract, per un totale di 18.546 frasi, e circa mezzo milione di occorrenze di parole e 36 tipi di etichette di classi diverse. Il *corpus* di GENIA contiene annotazioni complesse per istanze riguardanti entità coordinate, in casi come «erythroid, myeloid and lymphod cell types». Noi abbiamo escluso dalla nostra analisi frasi contenenti esclusivamente istanze di entità complesse di questo tipo, e anche frasi contenenti più di 100 parole. Il numero finale di frasi utilizzate nei nostri esperimenti è di 18.333 (484.005 occorrenze, 91.387

entità). Molte etichette hanno una complessa struttura gerarchizzata, per esempio «[[IL-2 gene]DNA expression]Other_name». In questo studio ci concentriamo solo sulle etichette più interne, tuttavia le etichette esterne contengono certamente informazioni utili e dovrebbero essere utilizzate in futuro. Il buon uso delle etichette esterne dipende anche dallo sviluppo, in GENIA, della classe «Other_name», di cui si parlerà più sotto nella sezione sperimentale.

Un limite potenziale di GENIA è che definisce un numero relativamente piccolo di concetti biologici. La granularità dei concetti può risultare generica, e non distinguere fra sottoinsiemi importanti delle classi definite. A causa di ciò, alcune relazioni si adattano bene ad alcuni sottoinsiemi di una classe ma non altrettanto bene alla totalità della classe. Per esempio, il concetto

«DNA_domain_or_region» può etichettare sequenze di geni, ma anche promotori e geni. Anche se promotori e geni sono anche essi frammenti di sequenze, questi si riferiscono anche a descrizioni funzionali di sequenze. Per questo motivo, affermazioni diverse possono essere fatte riguardo a queste istanze di «DNA_domain_or_region». La relazione «DNA_domain_or_region encode Protein_molecule» ha senso se riguarda geni, ma non necessariamente se riguarda sequenze, a seconda della loro funzione. D'altra parte, qualsiasi risorsa destinata a supportare applicazioni di trattamento automatico del linguaggio è necessariamente limitata. Se definisse classi troppo specifiche l'ontologia sarebbe molto più complicata da usare automaticamente, e molto più onerosa da sviluppare. Rispetto a queste considerazioni, GENIA ha il vantaggio, cruciale, di fornire non solo un'ontologia ma anche un *corpus* esaustivamente annotato con le relative etichette. Questo *corpus* può essere usato per il training di annotatori automatici (cfr. KAZAMA *et al.* 2002).

3.2 Relazioni come dipendenze grammaticali

Una volta annotato il *corpus* con le classi dell'ontologia, il nostro metodo prevede l'analisi sintattica di tutte le frasi che lo compongono. Il *corpus* delle frasi di GENIA è a questo scopo analizzato sintatticamente con un *parser* statistico per la lingua inglese (CHARNIAK 2000). Poiché il nostro obiettivo è identificare relazioni fra entità, è desiderabile che le entità, nelle frasi, non finiscano spezzettate in sintagmi diversi. Questo fatto può verificarsi perché nomi di geni, proteine, ecc. possono essere composti da molti termini e avere una struttura morfologica complessa che include numeri, abbreviazioni e elementi di punteggiatura. Questo fatto può causare errori nelle strutture sintattiche generate dal *parser*.

Per evitare tali problemi, nelle frasi che vengono passate in input al parser, le occorrenze di classi di GENIA sono sostituite con le etichette delle classi ontologiche. In luogo della frase originale dell'esempio precedente (2) il parser analizza la frase seguente, le sostituzioni sono evidenziate in corsivo:

(3) Untreated *Cell_line* expressed *RNA_family_or_group* for the *Protein_molecule*.

Il risultato può anche essere osservato nella figura 3, che illustra come la struttura dei sintagmi nominali risulti notevolmente semplificata. Gli alberi sintattici ottenuti in questo modo sono più semplici e contengono meno errori. Nonostante le etichette possano essere parole sconosciute al parser, queste causano meno problemi dei complessi nomi delle istanze di categorie biologiche.

Per ogni sintagma è individuata la relativa 'testa': la parola da cui deriva la categoria sintattica del sintagma⁸; vedi anche fig. 3. Ogni parola nella frase ha un elemento governante, da cui dipende sintatticamente, il quale coincide con la testa di sintagma più vicina, diversa da essa stessa⁹. Ogni parola ha un solo e unico governante, e una volta identificate queste relazioni si ottiene un 'grafo di dipendenze', come esemplificato nella figura 3, dove «*Cell_line*» è governato/dipende da «*express*», mentre «*Protein_molecule*» dipende dalla preposizione «*for*», ecc.

A questo punto è possibile formalizzare la nozione, vaga fin qui, di relazione semantica. È da notare che questo tipo di rappresentazione, basata su grafi dipendenze, è stata usata anche per la ricognizione di frasi che sono rispettive parafrasi una dell'altra (LIN-PANTEL 2001). Questo problema interessa anche il nostro metodo, e sarà discusso ancora più sotto, nelle conclusioni.

Una relazione r tra due entità c_i e c_j , in un albero sintattico, è il cam-

⁸ Normalmente un nome per un sintagma nominale, un verbo per un sintagma verbale, ecc.

⁹ Le parole che costituiscono teste di sintagma che occorrono in forme flesse, vengono semplificate morfologicamente, cioè riportate alla forma morfologica di base (per esempio *children* = *child*, *went* = *go*, ecc.).

mino tra c_i e c_j , che segue le dipendenze definite dal grafo costruito sull'albero. Nella figura 3 il cammino tra «Cell_line» e «Protein_molecule» è « $\leftarrow$ express $\rightarrow$ for $\rightarrow$ ». Esiste un cammino tra ogni coppia di entità presenti in una frase. Inoltre, ogni cammino può essere considerato da due direzioni diverse, a seconda di quale argomento si considera per primo. In altri termini, ogni cammino da A a B ha un cammino inverso da B a A. In questo modo, nel *corpus* analizzato sintatticamente di GENIA, sono state individuate 172.446 differenti relazioni. Per facilitare l'interpretazione delle relazioni estratte dal *corpus*, e limitarne il numero in modo che sia fattibile l'ispezione delle relazioni selezionate, ci siamo limitati a considerare un sottoinsieme di tutti le relazioni estratte. In particolare, ci limitiamo a cammini fra c_i e c_j dove $j > i$, e l'elemento primario della relazione, quello che non dipende da nessun altro elemento, è un verbo, identificato con v . Inoltre, imponiamo la condizione che c_i sia governato da v sotto un sintagma di tipo S, cioè una frase completa. Quindi si tratta del soggetto di superficie del verbo, come nel caso di «Cell_line» in figura 3. Infine, una delle seguenti condizioni deve valere per il secondo argomento della relazione c_j :

C.1 – c_j è governato da v all'interno di un sintagma verbale (VP, c_j è il complemento oggetto di v , identificato con la sigla DIR_OBJ); per esempio «RNA_family_or_group» in figura 3;

C.2 – c_j è governato da v sotto un sintagma preposizionale (PP, c_j è il complemento indiretto di v , IND_OBJ); per esempio «Protein_molecule» nella figura 3;

C.3 – c_j è governato dal nome complemento oggetto di v (quindi specifica il complemento oggetto, DIR_OBJ_MOD); per esempio «Virus» nel contesto «... influence Virus replication»;

C.4 – c_j è governato dal nome complemento indiretto di v (specifica il complemento indiretto, IND_OBJ_MOD); per esempio «Protein_molecule» in «... was induced by Protein_molecule stimulation»;

C.5 – c_j è governato da una preposizione che modifica il complemento oggetto di v (DIR_OBJ_MOD_PP); per esempio «Protein_molecule» in «... induce overproduction of Protein_molecule»;

C.6 – c_j è governato da una preposizione che modifica il complemento indiretto di v (IND_OBJ_MOD_PP); per esempio «Lipid» in «... transcribed upon activation with Lipid».

Dalla frase dell'esempio (2) vengono quindi identificate due relazioni candidate: «SUBJ $\leftarrow$ express $\rightarrow$ DIR_OBJ», tra «Cell_line» e «RNA_family_or_group», e «SUBJ $\leftarrow$ express $\rightarrow$ for $\rightarrow$ IND_OBJ», tra «Cell_line» e «Protein_molecule». In totale, seguendo questa procedura, sono emerse 7.189 istanze di relazioni candidate, distribuite come segue:

Tipo	Frequenza	Frequenza relativa
SUBJ-R-DIR_OBJ	1.746	0,243
SUBJ-R-IND_OBJ	1.572	0,219
SUBJ-R-DIR_OBJ_MOD_PP	1.156	0,161
SUBJ-R-DIR_OBJ_MOD	943	0,131
SUBJ-R-IND_OBJ_MOD_PP	911	0,127
SUBJ-R-IND_OBJ_MOD	861	0,120

In totale, queste relazioni definiscono 485 tipi: 3.573 tipi di relazione e 5.606 tipi di coppie di argomenti.

3.3 Prima fase: selezione di relazioni

Definiamo con *A* una coppia ordinata di classi definite in GENIA; per esempio, *A* = (Protein_domain,DNA_domain_or_region), e con *B* una relazione, definita da un grafo di dipendenze come descritto in precedenza; per esempio, *B* = «SUBJ $\leftarrow$ bind $\rightarrow$ DIR_OBJ». Per comodità ci riferiamo a una coppia *AB* anche con il termine 'pattern'. Il nostro obiettivo è trovare un modo per selezionare relazioni (*B*) fortemente associate con coppie di argomenti (*A*). In altre parole vogliamo identificare pattern *AB* che esprimano interazioni significative nel dominio biologico-molecolare. Questo problema ha una formulazione simile a quello di trovare 'collocazioni' lessicali, cioè parole complesse costituite da più lemmi, come «hot spot», «fast food» e «no global», le quali formano espressioni idiomatiche. Il modo più semplice per identificare collocazioni si basa sulla frequenza in un *corpus*. Coppie che occorrono insieme spesso sono possibili collocazioni, o pattern nel nostro caso. Coppie *AB* possono occorrere spesso insieme semplicemente perché i termini *A* e/o *B* sono molto frequenti; per esempio «SUBJ $\leftarrow$ induce $\rightarrow$ DIR_OBJ» è una relazione molto frequente che appare insieme a 37 coppie di argomenti diverse. L'alta frequenza perciò può essere accidentale. Inoltre, questo metodo non offre una soluzione soddisfacente al problema di definire una soglia di frequenza minima, che discrimini pattern *AB* da selezionare, per essere inclusi nell'ontologia. Per questo motivo, seguendo metodi noti per la scoperta di collocazioni lessicali (DUNNING 1993), adottiamo, in luogo della frequenza, un semplice approccio statistico.

Una buona soluzione al problema di selezionare solo le coppie AB rilevanti è quella di eseguire un test statistico, per verificare se A e B occorrono insieme più frequentemente di quello che ci si aspetterebbe in base al semplice caso. Il metodo si basa sulla formulazione di un'ipotesi nulla', H_0 , in base alla quale A e B non occorrono insieme più frequentemente che a caso. In altri termini $P(AB) = P(A)P(B)$. Utilizzando statistiche estratte dal *corpus* di GENIA calcoliamo la probabilità $P(AB)$, sotto l'ipotesi H_0 . Se non è possibile distinguere la distribuzione empirica di A e B (la distribuzione basata sulle frequenze del *corpus*) dalla distribuzione predetta dal modello di H_0 , allora la coppia AB viene rigettata. Altrimenti la relazione B va ad incrementare l'insieme di nuove relazioni fra le classi in A. Per identificare le coppie AB rilevanti utilizziamo un test 'chi-square'. Per ogni coppia AB, creiamo una tavola di contingenza per le frequenze dei casi AB, $A \neg B$, $\neg AB$, $\neg A \neg B$. Per esempio, nel caso in cui $A = (\text{Protein_molecule}, \text{DNA_domain_or_region})$, e $B = \langle \text{SUBJ} \leftarrow \text{bind} \rightarrow \text{DIR_OBJ} \rangle$, la tavola è la seguente (6, 161; 24, 6,988). Insieme al test, utilizziamo la seguente statistica¹⁰:

$$(4) G^2 = 2n \sum_{i,j} o_{ij} \log(o_{ij}/e_{ij})$$

dove i e j indicizzano, rispettivamente, le righe e le colonne della tavola di contingenza, le osservazioni (o_{ij}), e le frequenze predette da H_0 (e_{ij}) sono calcolate dai margini della tavola di contingenza. Nell'esempio precedente G^2 è uguale a 16,43, quindi ha un valore maggiore del valore critico 7,88 per $\alpha = 0,005$, perciò B è accettata come relazione rilevante per la coppia A. La tabella seguente mostra le tre coppie A che ottengono i valori maggiori di G^2 per la relazione B. Dai risultati del test statistico emerge una forte associazione di B con le prime due coppie A, mentre la terza non supera il valore critico di α . In altri termini, c'è una forte evidenza del fatto che entità del tipo proteina, nel ruolo di soggetto di «bind» sono fortemente associate con elementi di DNA, una conclusione ragionevole dal punto di vista biologico:

¹⁰ Una statistica *log-likelihood chi-squared*. Dunning (1993) sostiene che G^2 è più appropriata della statistica X^2 di Pearson, in presenza di dati scarsi, nel nostro caso entrambe producono risultati comparabili.

B = SUBJ $\leftarrow$ bind $\rightarrow$ DIR_OBJ		
A	G ²	Significativo
Protein_domain - DNA_domain_or_region	16,43	SI
Protein_family_or_group - DNA_domain_or_region	13,67	SI
Virus - Protein_molecule	7,84	NO

In questo lavoro abbiamo utilizzato un valore prefissato di α di 0,005. In generale, α definisce un parametro libero, che potrebbe essere determinato in modo empirico così da massimizzare una funzione oggettiva esplicita. In questo caso noi ci limitiamo all'opzione più semplice di scegliere un valore conservativo, poiché il nostro obiettivo primario è massimizzare la precisione del metodo, quindi la selezione di pattern attendibili. Inoltre, come ulteriore filtro *a priori*, ignoriamo coppie AB che occorrono insieme meno di 2 volte, e coppie A e relazioni B che occorrono individualmente meno di 4 volte. In generale, in questo modo si definisce un insieme di 487 candidati AB, di questi 287 (58.6%) raggiungono un valore di G² superiore ad α e sono selezionati come attendibili.

3.4 Seconda fase: generalizzazione di relazioni

Come nell'esempio di sopra relativo a «bind», alcune relazioni possono avere argomenti caratterizzati da classi ontologiche simili. Nella tabella, in entrambe le relazioni selezionate, il complemento oggetto è «DNA_domain_or_region» mentre il soggetto è una qualche forma di proteina. Queste similarità possono suggerire la possibilità che in realtà non ci troviamo di fronte a due diversi pattern, ma si tratta di un pattern più generale. Vale a dire riguardante una classe superordinata degli argomenti in A, di cui differenti istanze vengono identificate. In questi casi sarebbe desiderabile, se possibile, apprendere il pattern più generale. Un pattern generalizzato si applica a più entità e definisce un modello concettuale più compatto. Per esempio, «Protein $\leftarrow$ bind $\rightarrow$ DNA», in luogo dei due pattern nella tabella che superano il test statistico. A questo punto si può notare che il task di individuare il livello di astrazione più adatto per un argomento di una relazione è un caso particolare di un altro problema di apprendimento semantico ampiamente studiato. Il problema di apprendere le «preferenze semantiche» di verbi e predicati; cioè le preferenze selettive che i predicati definiscono per i propri argomenti. Come nel caso del verbo «mangiare», che seleziona complementi oggetto che appartengono alla classe dei cibi, mentre «cinguettare» seleziona soggetti che appartengono alla classe degli uccelli, ecc. Diversi metodi di apprendimento automatico sono stati proposti per questo problema, vedi (LIGHT-GREIFF 2002). Noi

adottiamo qui il metodo di Clark e Weir (2002), uno dei più accurati, semplici ed efficienti. La struttura gerarchica di supporto per la generalizzazione è data dalla tassonomia dell'ontologia di GENIA, di cui la figura 2 illustra una porzione.

Clark e Weir (2002) propongono un algoritmo per l'apprendimento di preferenze selettive basato su una funzione, denominata $top(c,r,s)$. Questa funzione, adattando la terminologia al nostro caso, prende come input una relazione r , una classe c e un argomento s e restituisce una classe c' , che può essere c stessa, oppure una classe superordinata di c nella tassonomia di GENIA. La classe restituita è quella che determina la migliore generalizzazione per la probabilità condizionale $P(r|c,s)$. Anche questo metodo utilizza un test statistico chi-square per verificare se la distribuzione $P(r|c,s)$ è significativamente diversa da $P(r|c',s)$, dove c' è la classe immediatamente superordinata di c . Se la differenza non è significativa, allora si presume che $P(r|c',s)$ sia una buona approssimazione. Questo viene interpretato come evidenza del fatto che la relazione (r,s) vale anche per c' , oltre che per c . La procedura si applica iterativamente, finché non emerga una differenza significativa. A quel punto la classe corrente è identificata come la classe ottimale per la relazione r rispetto all'argomento s . In altre parole, la classe selezionata dalla relazione in quell'argomento. Le frequenze delle classi superordinate, necessarie per eseguire l'algoritmo, possono essere calcolate sommando per ogni classe superordinata le frequenze di tutti i suoi discendenti in GENIA. Nel caso dell'esempio della tabella l'algoritmo produce infatti la generalizzazione desiderata del pattern: «Protein $\leftarrow$ bind $\rightarrow$ DNA». La figura 4 mostra, in un esempio più consistente, l'insieme di pattern riguardanti la classe Virus, selezionate dal *corpus* di GENIA col nostro metodo.

4 Valutazione

La valutazione dell'output dell'algoritmo è stata eseguita manualmente da esperti di biologia molecolare, bioinformatica e ingegneria ontologica nel campo biomedico, con un'esperienza approfondita di GENIA. La valutazione biologica si focalizza principalmente sulla 'precisione' del metodo, più precisamente, la percentuale di relazioni selezionate dal sistema che esprimono interazioni corrette tra i due argomenti della relazione. Dal punto di vista ontologico analizziamo invece gli aspetti semantici delle relazioni apprese, in particolare il grado di consistenza con le classi di GENIA.

4.1 Valutazione biologica

L'output della prima fase di apprendimento è un insieme di 287 pattern AB, composti da una coppia ordinata di classi di GENIA e da una relazione semantica. Novantuno di queste relazioni, contenenti in uno o entrambi gli argomenti, la classe «Other_name» sono risultati non valutabili, ed escluse dalla valutazione. Il problema consiste nel fatto che la classe «Other_name» di GENIA costituisce una sorta di contenitore, a un livello molto generale, per tutti i concetti non ancora sviluppati nell'ontologia. La classe non ha alcuna organizzazione strutturale o partizione interna. Relazioni di questo tipo, per esempio *treat* (curare) possono risultare corrette per alcuni sottoinsiemi di argomenti etichettati con «Other_name», per esempio il complemento oggetto «inflammation», ma falsi per altri, per esempio il complemento oggetto «gene expression». Dei rimanenti 196 pattern AB 76,5% (150) risultano corretti, in altre parole esprimono fatti biologici veri, come «Protein_molecule $\leftarrow$ induce $\rightarrow$ phosphorylation $\rightarrow$ of $\rightarrow$ Protein_molecule», mentre il 23,5% (43) risultano incorretti; per esempio «Protein $\leftarrow$ inhibit $\rightarrow$ expression $\rightarrow$ of $\rightarrow$ Lipid». La valutazione manuale è stata eseguita anche attraverso l'ispezione esaustiva delle frasi originali, da dove le relazioni erano state estratte, per verificare il significato inteso originale e individuare tipi di errori informativi. Metà degli errori (22) dipende dal modo in cui il nostro sistema gestisce le coordinazioni, all'interno di sintagmi nominali, che ha come effetto che parte della struttura coordinata sia inclusa nella relazione. Per esempio, le prime due istanze di DNA nel sintagma nominale «DNA, DNA and DNA» sono governate dall'ultima istanza di DNA, la testa del sintagma. Nonostante siano tutte sintatticamente allo stesso livello. In questi casi, relazioni erronee come «DNA $\leftarrow$ bind $\rightarrow$ DNA $\rightarrow$ DNA» sono generate come candidate, oltre alla corretta «DNA $\leftarrow$ bind $\rightarrow$ DNA». Questi casi non sono esattamente degli errori. Poiché la struttura della coordinazione è coerente col modo in cui abbiamo definito le relazioni. La correzione di questo problema riguarda semplicemente lo sviluppo di un filtro adeguato per questi casi, oppure l'adozione di una rappresentazione più sofisticata delle coordinazioni. Altri 5 errori dipendono dal fatto che la classe «Other_name» occorre all'interno della relazione, rendendo impossibile la verifica del pattern. I rimanenti errori che non si riconducono a dei tipi specifici dipendono probabilmente da rumore derivato da dati insufficienti. Questo problema può essere alleviato semplicemente utilizzando più dati. Infine, notiamo che, nonostante GENIA sia stata sviluppata per lo studio delle interazioni cellulari, mancano dei concetti molto importanti in tale dominio come quello di *signaling pathway* (per-

corso segnalatore). Questo fatto determina che importanti relazioni non possano venire identificate, come nel caso seguente:

(5) An intact [TCR]_{Protein_molecule} signaling pathway is required for [p95vav]_{Protein_molecule} to function.

Da questa frase si può solo identificare un pattern, erroneo, come «Protein_molecule $\leftarrow$ is $\rightarrow$ required $\rightarrow$ for $\rightarrow$ Protein_molecule», poiché solamente «Protein_molecule» è disponibile come etichetta per TCR mentre ne manca una per «signaling pathway».

In uno studio simile al nostro Reinberger *et al.* (2004), presentano un sistema per estrarre triple soggetto-verbo-oggetto (SVO) in biologia molecolare, da un *corpus* di Medline. In particolare, il metodo seleziona un insieme di 165 pattern SVO che vengono poi valutati da un esperto biologo. Questo studio riporta una percentuale di accuratezza del 42% di relazioni corrette. Nonostante non sia possibile una comparazione diretta dei risultati dei due sistemi abbiamo cercato di investigare i motivi della differenza di prestazione fra i due approcci concentrandoci sulle differenze strutturali dei due sistemi. Il metodo presentato in (REINBERGER *et al.* 2004) differisce dal nostro su tre punti importanti. Primo, le relazioni sono estratte tra i nomi che costituiscono le teste dei sintagmi nominali, e non fra entità dopo una fase di named entity recognition come nel nostro caso. Secondo, il metodo utilizza un parser superficiale (uno *shallow parser*), piuttosto che un parser completo. Infine, la selezione delle relazioni rilevanti è eseguita semplicemente in base alla frequenza delle triple. Le triple che occorrono meno di un certo numero di volte vengono eliminate, mentre il nostro metodo utilizza un test di ipotesi statistico. Le differenti prestazioni possono dipendere da ogni possibile combinazione di questi elementi, e forse altri. Tuttavia, possiamo formulare alcune ipotesi sull'impatto di questi elementi considerati individualmente. La scelta di filtrare relazioni in base alla frequenza semplice, non è secondo noi, un fattore che può determinare una differenza simile. La frequenza è di per se un filtro accettabile e naturale, il problema maggiore è che non permette di definire una soglia critica in un modo motivato. Per quanto riguarda il parsing, Punyakanok *et al.* (2005), in un'analisi dell'impatto di diverse tecnologie di parsing su un task di classificazione di ruoli grammaticali, notano una maggiore performance (circa del 5%) di un modello basato sul full parsing piuttosto che sullo shallow parsing. Questo elemento quindi potrebbe avere avuto un impatto significativo, anche se probabilmente non spiega tutta la differenza. Secondo noi, l'e-

lemento cruciale è dato dalla pre-individuazione delle entità (nel nostro caso già annotate, o tramite NER). Poiché questo permette l'accumulazione di statistiche migliori, e semplifica allo stesso tempo il compito del parser e la qualità delle dipendenze estratte.

Infine, per quanto riguarda l'apprendimento di relazioni generalizzate nella seconda fase del sistema la valutazione è stata eseguita nel modo seguente. Prima di tutto, abbiamo rimosso tutte le relazioni che riguardavano, in uno o tutti e due gli argomenti, la classe «Other_name» (40 su 153), poiché questa classe non ha super né sub-ordinati. Delle rimanenti 113 coppie AB, con argomenti generalizzati, abbiamo valutato quali delle generalizzazioni fossero corrette. In totale 60 (53,1%) relazioni generalizzate sono risultate corrette. Per esempio, «Protein_molecule $\leftarrow$ induce $\rightarrow$ phosphorylation $\rightarrow$ of $\rightarrow$ Amino_acid_monomer» è sostituita da «Protein $\leftarrow$ induce $\rightarrow$ phosphorylation $\rightarrow$ of $\rightarrow$ Amino_acid_monomer». In generale, questa fase di apprendimento risulta considerevolmente meno precisa della prima. Il motivo dipende in larga parte dalla struttura tassonomica che ha il compito di supportare tali generalizzazioni. Come già accennato, la tassonomia di GENIA non è una gerarchia Is-a; per esempio, «DNA_substructure» non è un tipo di «DNA» e «Protein» non è un tipo di «Amino_acid». Per fare in modo che generalizzazioni del tipo selettivo siano inferite correttamente, quindi è necessario che la struttura tassonomica sia chiarita e le relazioni Is-a distinte dalle altre generiche relazioni strutturali/tassonomiche.

4.2 Valutazione ontologica

L'insieme di pattern AB giudicati corretti nella fase di valutazione biologica vengono ora valutati dal punto di vista ontologico, in quanto candidati ad arricchire l'ontologia di GENIA. Allo scopo di svolgere questa analisi, GENIA, incluse le nuove relazioni, è stata compilata in un linguaggio, OWL (Ontology Web Language, McGUINNESS-VAN HARMELLEN 2004), per esprimere in modo esplicito il significato degli elementi dell'ontologia ed esaminare le sue proprietà con strumenti di ingegneria ontologia. Ignorando «Other_name», la tassonomia di GENIA è caratterizzata da due nodi che sussumono tutte le altre classi più specifiche, definendo una partizione delle classi di GENIA in due insiemi disgiunti: *Source* (organismi, cellule, ecc.) e *Substance* (DNA, proteine, amino acidi, ecc.). È importante notare che questa partizione deriva da un principio di design strutturale dell'ontologia. Il nostro obiettivo principale nella fase di valutazione ontologica è di verificare il grado in cui le relazioni apprese sono coerenti con questo principio fondamentale di GENIA.

Allo scopo di analizzare le relazioni, in primo luogo, mappiamo le due superclassi *Source* e *Substance* a classi equivalenti di un'altra ontologia più generale, sviluppata indipendentemente. Sarebbe desiderabile che questo allineamento avesse luogo rispetto a un'ontologia dello stesso dominio, come TAMBIS (STEVENSON *et al.* 2000). Purtroppo TAMBIS non si basa sugli stessi principi strutturali di GENIA e distribuisce in altro modo le sottoclassi di *Source* e *Substance*, le quali inoltre non coprono gli stessi domini (per esempio *Substance* in TAMBIS non comprende proteine e acidi nucleici). In GENIA i subordinati di *Substance* sono classificati in base alle loro caratteristiche chimiche piuttosto che in base alle loro funzioni biologiche. Mentre i subordinati di *Source* sono organismi, o parti di organismi, biologici dove le 'sostanze' sono collocate, o le relazioni che li coinvolgono hanno luogo. Questa distinzione presuppone uno schema in cui il livello ontologico biologico è sovrapposto al livello chimico. Questa caratteristica di GENIA la rende adatta ad essere allineata con un'altra ontologia, DOLCE-lite-plus (DLP, <<http://dolce.semantic-web.org>>), una traduzione semplificata dell'ontologia fondazionale DOLCE (GANGEMI *et al.* 2003). DLP definisce una distinzione compatibile con GENIA tra oggetti chimici e biologici. DLP è costituita da circa 200 classi, 150 relazioni, e 500 assiomi, ed è stata utilizzata in molteplici domini inclusa la biomedicina (ŠARIĆ *et al.* 2004a). *Source* e *Substance* vengono così allineate alle classi biologiche e chimiche di DLP.

Il risultato dell'analisi ontologica è il seguente. Ci sono 78 tipi di relazioni diverse tra i 150 pattern correttamente selezionati. Il 58% di questi (45) occorre solo con una coppia di argomenti, mentre 33 sono relazioni polisemiche che occorrono con molteplici coppie. Dato che le superclassi *Source* e *Substance* definiscono due partizioni disgiunte dell'ontologia, abbiamo verificato se ci sono relazioni polisemiche che occorrono, nello stesso argomento, con classi che sono sottoclassi di entrambi *Source* e *Substance*. Tali relazioni non implicano necessariamente inconsistenze logiche, tuttavia andrebbero identificate e analizzate con attenzione poiché, in un certo senso, suggeriscono l'emergere dai dati testuali di classi che violano la supposta disgiunzione fra le superclassi. Solamente quattro relazioni di questo tipo emergono dai dati. Un esempio è «encode», il cui soggetto può essere sia «Virus» (*Source*) che «DNA» (*Substance*). In biologia, il DNA codifica una proteina, ma a volte i biologi usano il termine in modo metonimico. Dicendo che un virus codifica una proteina, ciò che realmente vogliono dire è che è il genoma del virus che codifica la proteina. È importante chiarire questi casi, laddove si voglia associare una semantica esplicita alle relazioni. I pochi casi

di, apparenti, contraddizioni di questo tipo suggerisce che le relazioni emergenti dal *corpus* sono coerenti, al livello più generale, con i principi organizzativi di GENIA.

A un livello semantico più specifico troviamo che la composizione delle relazioni è la seguente: 54 (68%) rappresentano ‘eventi’, cioè codificano la concettualizzazione di processi chimici che hanno luogo in organismi. L’81% delle relazioni tra una superclasse e l’altra sono eventive, dando supporto quindi all’assunzione di GENIA che processi chimici biologicamente rilevanti coinvolgono sia un oggetto biologico che uno chimico. Le relazioni non-eventive hanno soprattutto un significato ‘strutturale’ (per esempio «Consist-of»), oppure locativo («located-in»), oppure anche epistemologico («Identified-in»).

5 Conclusioni

Questo lavoro presenta uno studio sul problema di apprendere relazioni semantiche da testi nel dominio della biologia molecolare. Abbiamo presentato un metodo, che non richiede alcuna supervisione, e lo abbiamo applicato al *corpus* e all’ontologia di GENIA. Il metodo si basa sulla rappresentazione delle relazioni come grafi di dipendenza sintattica, tra coppie ordinate di classi di GENIA, identificate nel testo come *named entities*. Per selezionare relazioni rilevanti il metodo applica un test statistico ai pattern, composti da una coppia ordinata di classi e una relazione. Abbiamo anche mostrato che è anche possibile generalizzare gli argomenti delle relazioni utilizzando tecniche di apprendimento di ‘restrizioni semantiche’ di predicati. I risultati della valutazione del metodo dimostrano come questo sia caratterizzato da una alta precisione. Inoltre l’approccio è caratterizzato da semplicità concettuale e implementativa¹¹.

Altri aspetti interessanti possono esplorati in questa ricerca. Tra gli

¹¹ Questo articolo rielabora ed estende la ricerca descritta in (CIARAMITA *et al.* 2005), un lavoro svolto insieme ad Aldo Gangemi, Esther Ratsch, Jasmin Šarić e Isabel Rojas, che ringrazio per la collaborazione, insieme a tutti i membri del LOA.

altri, la valutazione del grado di copertura del metodo, e la sua utilità in applicazioni di estrazione di informazione. Quest'ultimo problema necessita lo sviluppo di metodi per individuare relazioni sinonimiche; per esempio, nel contesto delle interazioni fra proteine, «positively regulate» è sinonimo di «activate», «up-regulate», «derepress», «stimulate», ecc. La scelta della rappresentazione delle relazioni come dipendenze sintattiche ci permette di estendere il modello in questa direzione in modo naturale, dato che il problema di individuare sinonimi di relazioni è riconducibile al problema di individuare parafrasi che può essere affrontato con metodi statistici come nel caso di (LIN-PANTEL 2001) che utilizzano una rappresentazione simile alla nostra.

MASSIMILIANO CIARAMITA

Bibliografia

- ALMUHAREB-POESIO 2004 = A. ALMUHAREB-M. POESIO, *Attribute-Based and Value-Based Clustering: An Evaluation*, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, Barcellona 2004
- BERLAND-CHARNIAK 1999 = M. BERLAND-E. CHARNIAK, *Finding Parts in Very Large Corpora*, in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999)*, College Park, Maryland 1999
- CHARNIAK 2000 = E. CHARNIAK, *A Maximum-Entropy Inspired Parser*, in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, Hong Kong 2000
- CIARAMITA et al. 2005 = M. CIARAMITA-A. GANGEMI-E. RATSCH-J. ŠARIĆ-I. ROJAS, *Unsupervised Learning of Semantic Relations between Concepts of a Molecular Biology Ontology*, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2005)*, Edimburgh 2005.
- CLARK-WEIR 2002 = S. CLARK-D. WEIR, *Class-Based Probability Estimation Using a Semantic Hierarchy*, in «*Computational Linguistics*», XXVIII, 2, 2002
- CRAVEN-KUMLIEN 1999 = M. CRAVEN-J. KUMLIEN, *Constructing Biological Knowledge Bases by Extracting Information from Text Sources*, in *Proceedings of the 7th International Conference on Intelligent Systems for Molecu-*

- lar Biology (ISMB 1999), Heidelberg 1999
- DUNNING 1993 = T. DUNNING, *Accurate Methods for the Statistics of Surprise and Coincidence*, in «Computational Linguistics», XIX, 1993
- FRIEDMAN et al. 2001 = C. FRIEDMAN-P. KRA-H. YU-M. KTRAUTHAMER-A. RZHETSKY, *GENIES: A Natural-Language Processing System for the Extraction of Molecular Pathways from Journal Articles*, in «Bioinformatics», XVII, 2001
- GANGEMI et al. 2003 = A. GANGEMI-N. GUARINO-C. MASOLO-A. OLTRAMARI, *Sweetening WordNet with DOLCE*, in «AI Magazine», XXIV, 3, 2003
- GUARINO 1998 = N. GUARINO, *Formal Ontology in Information Systems*, in *Formal Ontology in Information Systems. Proceedings of FOIS '98*, ed by N. Guarino, Amsterdam, IOS Press 1998, , pp. 3-15
- HEARST 1992 = M. HEARST, *Automatic Acquisition of Hyponyms from Large Text Corpora*, in *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING 1992)*, College Park, Maryland 1992
- HEARST 1999 = M. HEARST, *Untangling Text Data Mining*, in *Proceedings the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999)*, (invited paper), 1999
- KAZAMA et al. 2002 = J. KAZAMA-T. MAKINO-Y. OHTA-J. TSUJII, *Tuning Support Vector Machines for Biomedical Named Entity Recognition*, in *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain (ACL 2002)*, Philadelphia 2002
- LIGHT-GREIFF 2002 = M. LIGHT-W. GREIFF, *Statistical Models for the Induction and Use of Selectional Preferences*, «Cognitive Science», LXXX-VII, 2002
- LIN-PANTEL 2001 = D. LIN-P. PANTEL, *DIRT – Discovery of Inference Rules from Text*, in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2001)*, San Francisco 2001
- MCGUINNESS-VAN HARMELEN 2004 = D. MCGUINNESS-F. VAN HARMELEN, *Owl Web Ontology Language Overview*, in *W3C Recommendations: <<http://www.w3c.org/TR/owl-features/>, 2004>*
- OHTA et al. 2002 = Y. OHTA-Y. TATEISI-J. KIM-H. MIMA-J. TSUJII, *The GENIA Corpus: An Annotated Corpus in the Molecular Biology Domain*, in *Proceedings of the Human Language Technology Conference (HLT 2002)*, San Diego 2002
- PANTEL-RAVICHANDRAN 2004 = P. PANTEL-D. RAVICHANDRAN, *Automatically Labelling Semantic Classes*, in *Proceedings of the Human Language Techno-*

- logy Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004), Boston 2004
- PUSTEJOWSKY et al. 2002 = J. PUSTEJOWSKY-J. CASTAÑO-B. ZHANG-B. COCHRAN-M. KOTECI, *Robust Relational Parsing over Biomedical Literature: Extracting Inhibit Relations*, in *Proceedings of the Pacific Symposium on Biocomputing (PSB 2002)*, Hawaii 2002
- PUNYAKANOK et al. 2005 = V. PUNYAKANOK-D. ROTH-W. YIH, *The Necessity of Syntactic Parsing for Semantic Role Labeling*, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2005)*, Edinburgh 2005, pp. 1117-1123
- RATSCH et al. 2003 = E. RATSCH-J. SCHULTZ-J. ŠARIĆ-P. CIMIANO-U. WITTIG-U. REYLE-I. ROJAS, *Developing a Protein Interactions Ontology*, in «Comparative and Functional Genomics», IV, 2003, pp. 85-89
- ROJAS et al. 2002 = I. ROJAS-L. BERNARDI-E. RATSCH-R. KANIA-U. WITTIG-J. ŠARIĆ, *A Database System for the Analysis of Biochemical Pathways*, in «Silico Biology» II, 7, 2002
- ŠARIĆ et al. 2004a = J. ŠARIĆ-A. GANGEMI-E. RATSCH-I. ROJAS, *Modelling Gene Expression*, in *Proceedings of the Workshop on Models and Metaphors from Biology to Bioinformatics Tools (NETTAB 2004)*, Camerino 2004
- ŠARIĆ et al. 2004b = J. ŠARIĆ-L.J. JENSEN-R. OUZOUNOVA-I. ROJAS-P. BORK, *Extraction of Regulatory Gene Expression Networks from PubMed*, in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcellona 2004
- STEVENS et al. 2000 = R. STEVENS-P. BAKER-S. BECHOFER-G. NG-A. JACOBI-N.W. PATON-C.A. GOBLE-A. BRASS, *TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources*, in «Bioinformatics», XVI, 2, 2000
- SWANSON-SMALHEISER 1997 = D.R. SWANSON-N.R. SMALHEISER, *An Interactive System for Finding Complementary Literatures: A Stimulus to Scientific Discovery*, «Artificial Intelligence», XCI, 2, 1997

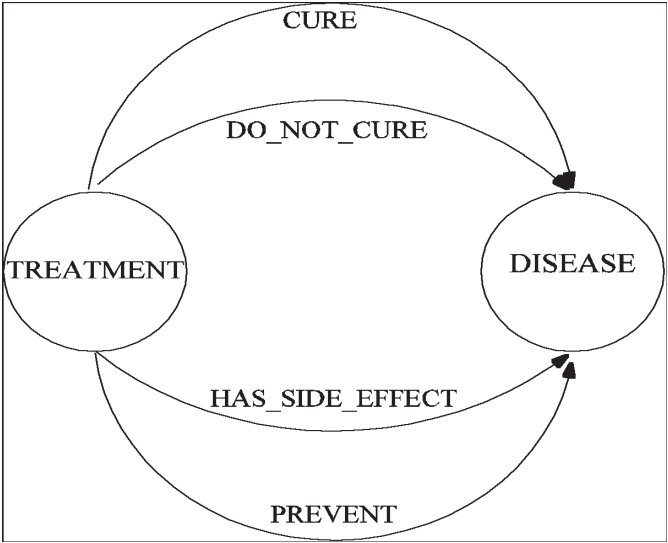


Fig. 1. Un modello schematico dell'ontologia utilizzata per l'estrazione di entità e relazioni in (Rosario-Hearst 2004). Le ellissi rappresentano le classi, le frecce (con direzione da causa a effetto) rappresentano le relazioni definite fra coppie di classi.

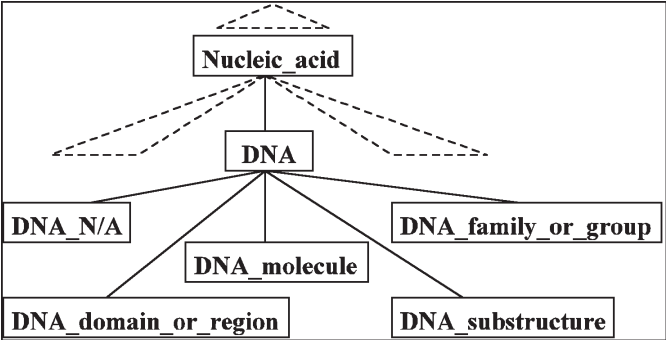


Fig. 2. Una frazione della tassonomia di GENIA riguardante la regione circostante la classe DNA.

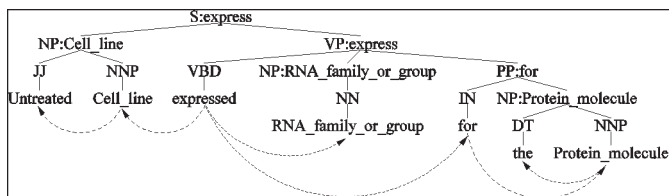


Fig. 3: Albero sintattico generato dal parser per la frase «Untreated Cell_line expressed RNA_family_or_group for the Protein_molecule». I sintagmi sono annotati con le rispettive teste (semplificate morfologicamente), le frecce indicano le relazioni di dipendenza (il grafo di dipendenze).

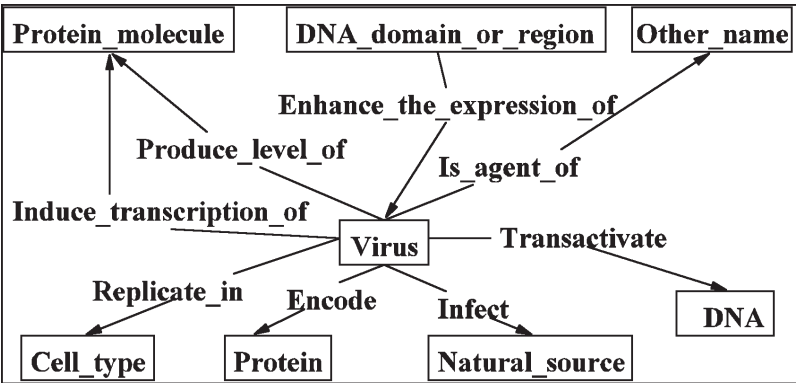


Fig. 4. La classe «Virus» di GENIA con le relazioni rilevanti, e le rispettive classi correlate, estratte dal corpus, gli argomenti delle relazioni «Encode», «Infect» e «Transactivate» sono classi generalizzate.

Etica di Internet

1 Premessa

L'applicazione delle nuove tecnologie all'ambito della comunicazione ha provocato, com'è ormai esperienza diffusa, non solo un potenziamento e una trasformazione dei *media* tradizionali, ma anche la nascita di *new media*, capaci di utilizzare al meglio le opportunità offerte dai progressi tecnologici. Sbaglierebbe tuttavia chi ritenesse questi mezzi di comunicazione solamente, appunto, dei 'mezzi': semplici strumenti da usare per scopi ad essi estrinseci. Giacché è pure esperienza quotidiana il fatto che essi non sono in sé qualcosa di neutro, di neutrale, che acquista valore e significato a seconda del fine per il cui raggiungimento sono impiegati.

Infatti le tecniche della comunicazione, vecchie o nuove che siano, da sempre interagiscono con il pensiero e con la prassi dell'uomo, modificando il suo rapporto con gli altri uomini e con il mondo. Basti pensare alle trasformazioni che nella storia si sono accompagnate al passaggio dalla dimensione dell'oralità a quella della scrittura, o dalla civiltà della scrittura a quella dell'immagine¹. Nel caso dei *new media*, però, l'impatto sui comportamenti umani delle nuove tecnologie, in quanto applicate all'ambito della comunicazione, risulta davvero di grande portata, tanto che da più parti si è parlato di una nuova «mutazione antropologica» di cui esse sarebbero responsabili.

Tutto ciò significa che le domande dell'etica, in questo caso, non sono solamente relative ai modi in cui tali mezzi di comunicazione possono essere utilizzati. Esse debbono riguardare, anche e soprattutto, la loro

¹ Sull'argomento, ad esempio, si veda l'ormai classico testo di D. DE KERCKHOVE, *La civilizzazione video-cristiana*, trad. it. a cura di C. Peltier/Shake, Milano, Feltrinelli 1995 (1990).

configurazione, gli effetti delle loro procedure, i cambiamenti che essi possono esercitare sull'uomo e sul mondo. In altre parole: i comportamenti che possiamo adottare, da un lato, 'rispetto alle nuove forme di comunicazione' e, dall'altro, 'all'interno di esse'. A ben vedere, sono proprio questi due piani ad essere stati studiati, pur a diversi livelli di approfondimento, dai contributi che, a partire dalla metà degli anni Ottanta, hanno in particolare discusso il tema «*Computers & Ethics*»².

In questo scritto ci limiteremo a prendere in esame, fra le nuove tecnologie multimediali, soprattutto Internet, la «rete delle reti», cercando di fare chiarezza su alcune delle questioni di carattere etico implicate da questo fenomeno e dal suo progressivo diffondersi. Lo faremo, conformemente a quanto appena detto, seguendo due linee interpretative. Esse riguardano, schematicamente:

- Le questioni relative all'impatto di Internet sulla società, sul mondo in cui viviamo, sul nostro modo di pensare, indipendentemente dal fatto che, concretamente, abbiamo o meno a che fare con Internet. Infatti,

² Il primo testo che cerca di fare il punto su tale questione – sia bandendo un concorso relativo a *papers* sull'argomento e pubblicando i testi giudicati migliori, sia proponendo una bibliografia ragionata e recensendo i libri fino a quel momento apparsi riguardo al nostro tema – è il fascicolo 16, ottobre 1985, della rivista «*Metaphilosophy*» (Basil Blackwell, Oxford and New York), a cura di Terrell Ward Bynum, appunto dedicato a «*Computer & Ethics*». Si vedano oggi, per un'introduzione all'argomento, almeno i seguenti contributi: D. MASSARO-A. GROTTI, *Il filo di Sofia. Etica, comunicazione e strategie conoscitive nell'epoca di internet*, Milano, Bollati Boringhieri 2000; D. LANGFORD, *Internet Ethics*, London, Macmillan 2000; D. JOHNSON, *Computer Ethics*, Upper Saddle River (NJ), Prentice Hall 2001²; D.M. HESTER-P.J. FORD, *Computers and Ethics in the CyberAge*, Upper Saddle River (NJ), Prentice Hall 2001; D. WALTON, *Internet ... e poi? Teoria critica dei nuovi media*, trad. it. di C. Marullo Reedtz, Bari, Dedalo 2001; *Internet e l'esperienza religiosa in rete*, a cura di P. Aroldi, B. Scifo, Milano, Vita & Pensiero 2002; J.D. BOLTER, *Lo spazio dello scrivere. Computer, ipertesto e la ri-mediazione della stampa*, trad. it. di S. Galli, Milano, Vita & Pensiero 2002; P. D'ALESSANDRO, *Critica della ragione telematica*, Milano, LED 2002; C.J. HAMELINK, *The Ethics of Cyberspace*, London, Sage 2002; A. MARTURANO, *The Role of Metaethics and the Future of Computer Ethics*, in «*Ethics and Information Technology*», IV, 2002, pp. 71-78; E. ROOKSBY, *E-mail and Ethics: Style and Ethical Relations in Computer-Mediated Communications*, Routledge, London 2002; G. ROMANO, *La città che non c'è. L'internet, frontiera di uomini*, Roma, Edizioni Lavoro 2004; *Etica del Virtuale*, a cura di A. Fabris, Milano, Vita e Pensiero 2007.

anche se non siamo programmatori, *webmasters*, o anche semplici utilizzatori della rete – «navigatori», per usare una metafora corrente, o più precisamente *netsurfers* –, facciamo comunque parte di una realtà su cui ormai Internet ha inciso profondamente. E dunque siamo chiamati in qualche modo a confrontarci con essa.

Un tale ambito di ricerca lo possiamo definire, con un'espressione standard, il filone relativo all'*etica di Internet*. Rientrano qui le riflessioni sulle conseguenze del divario tecnologico (il cosiddetto *digital divide*), sugli eventuali usi militari di Internet³ e, più in generale, sull'impatto sociale delle nuove tecnologie. E la responsabilità che è qui in gioco riguarda anzitutto i ricercatori e gli operatori che hanno a che fare con Internet, ma non esclude i suoi semplici fruitori, grazie ai quali soltanto la «rete delle reti» può continuare ad espandersi.

- In parallelo debbono essere considerati più da vicino i problemi di comportamento relativi all'uso della rete stessa. Una tale indagine è quella che riguarda, più propriamente, ciò che possiamo chiamare *l'etica in Internet*. Alla necessità di stabilire tali criteri di condotta si è cercato finora di dare risposta, per lo più, attraverso lo strumento dei codici e dei protocolli d'intesa fra gli operatori del settore⁴. Ma non basta muoversi ad un livello giuridico-normativo. Così come non basta dare semplici indicazioni di comportamento, o esortare a un uso corretto della rete. L'etica non si risolve, infatti, nell'etichetta. Oltre ad elaborare questi criteri del corretto uso di Internet, com'è stato da più parti fatto, bisogna allora cercar di dare adeguata giustificazione ai comportamenti che si ritiene debbano essere assunti. E anche questo è uno dei compiti dell'*«etica in Internet»*.

2 Le trasformazioni prodotte dalla rete e l'etica di Internet

³ Non si dimentichi che il primo progetto di Internet, il *network* di computer ARPANET, è stato proposto da Rand Corporation al Ministero della Difesa americano per costruire un sistema di comunicazioni militari che fosse invulnerabile a un attacco nucleare.

⁴ Si veda, per un primo approccio, i documenti contenuti nel CD in allegato a volume di *Guida alle etiche della comunicazione. Ricerche, documenti, codici*, a cura A. Fabris, Pisa, Edizioni ETS 2004.

Se vogliamo delineare, anzitutto, un'etica di Internet dobbiamo aver chiari gli specifici cambiamenti, nella mentalità e nei modi di agire, che la presenza della rete comporta. A questo proposito è bene accennare almeno, da un punto di vista filosofico, alle trasformazioni concettuali implicite nella pratica della rete e nell'interazione con essa. Queste soprattutto riguardano le nozioni di 'reale' e di 'virtuale', di 'spazio' e di 'tempo', di 'interno' e di 'esterno', nonché i possibili mutamenti, in virtù delle nuove tecnologie, del nostro rapporto con il corpo, e l'imporsi di quei modelli alternativi di espressione e di argomentazione che esse rendono possibile. Discutiamo allora, brevemente, una tale situazione per meglio definire le condizioni del nostro agire «nell'età della rete».

2.1 Com'è noto, uno dei concetti chiave che emergono in questa epoca è il concetto di 'virtuale'. Si parla di «tastiera virtuale», di «processi virtuali», addirittura di «realtà virtuale». Come si vede anche da quest'ultima espressione, 'virtuale' non è una nozione contrapposta a quella di 'reale'. Si tratta invece di un concetto più articolato e complesso, nella misura in cui racchiude e collega fra loro gli aspetti della 'possibilità' – nozione che indica una maggiore ampiezza e ricchezza di determinazioni rispetto a ciò che risulta effettivamente esistente –, della 'potenzialità' – termine che esprime la condizione in cui qualcosa si trova prima di attuarsi davvero: la pre-condizione, cioè, che ne rende possibile il compimento – e della 'potenza' – vocabolo che qui va inteso nel senso del potere, che qualcosa ha in sé, di compiere uno specifico atto o di realizzarsi in un certo modo. Conformemente a tali caratteri, 'virtuale' è appunto ciò che possiede la letteralmente la *virtus*, la «virtù», la capacità di fare qualcosa, e specialmente di farsi come qualcosa. In altre parole: non solo ciò che ha la potenzialità di realizzarsi in maniera conforme alla sua natura, ma soprattutto ciò che ha il potere di diventare ciò che esso è già in potenza, ciò che ha in sé la virtù per attuarsi dinamicamente secondo le varie possibilità in esso inscritte⁵.

Ecco perché, come mostra anche l'espressione «realtà virtuale» che

⁵ In questa lettura del virtuale mi discosto in parte dalle tesi espresse da P. LÉVY nel suo famoso testo *Il virtuale*, trad. it. di M. Colò e M. Di Sopra, Milano, Cortina 1997 (1995). Sull'argomento si vedano anche, almeno, i lavori di G. VENTIMIGLIA, *Ontologia ed etica del virtuale*, in «Teoria», 2004/1, pp. 119-47, e di R. DIODATO, *Estetica del virtuale*, Milano, Bruno Mondadori 2005.

abbiamo poc'anzi richiamato, non vi è affatto una contrapposizione tra 'reale' e 'virtuale'. Ecco perché il virtuale si configura davvero come un potenziamento della realtà: la messa in luce di quelle ulteriori possibilità dell'uomo, in lui insite, che le nuove tecnologie danno il potere di realizzare. E in questa dinamica è appunto implicito, tendenzialmente, anche un giudizio morale: giacché una tale realizzazione è appunto «virtù», comunque essa si configuri, nella misura in cui esprime la capacità di qualcosa di fare e di essere ciò che questo qualcosa può, tendenzialmente, fare ed essere. L'esercizio di tale capacità, potenziato o addirittura reso possibile dalla tecnologia, è considerato, insomma, già di per sé un bene.

Emerge così l'idea implicita per cui si ritiene eticamente giustificato, in ogni caso, l'esercizio della propria *virtus*. In ogni caso. Di modo che, per essere buoni, basta semplicemente «divenire ciò che già si è». E «buono», in questa prospettiva, sarebbe ogni mezzo che consente la propria realizzazione. Ma simili tesi, lungi dal poter essere immediatamente accettate, richiedono quanto meno di venire approfondite e discusse.

Insieme a questi problemi, però, se ne annuncia un altro, pur esso di grande importanza. Il fatto che il virtuale prenda, nel nostro mondo, sempre più spazio non significa solamente che esso aumenta la nostra capacità di concretizzare ciò che in passato rimaneva una mera possibilità, non indica unicamente un sempre più deciso processo di realizzazione, grazie alle tecnologie informatiche, di specifiche potenzialità. Come emerge dalla nostra esperienza quotidiana, infatti, vale anche il processo inverso: insieme alla 'realizzazione del virtuale' si ha oggi una crescente 'virtualizzazione del reale', una sua perdita di consistenza, una riduzione – potremmo dire – del suo spessore. Il reale non è più considerato come qualcosa che c'è già da sempre, che esiste e resiste al nostro agire. È invece il risultato delle nostre manipolazioni, della nostra incidenza su di esso. Ed è tale solo in quanto si mostra conforme alle nostre capacità, che finiscono, proprio grazie alle nuove tecnologie, per annullare la sua resistenza.

La realtà viene dunque a risolversi nel suo apparire. E l'apparire tende a trasformarsi in apparenza. Si chiude insomma il cerchio: da una parte, infatti, il virtuale mira ad affermare il proprio potere di realizzazione; dall'altra, in questa sua realizzazione, apre alla realtà nuovi scenari possibili. Ma ciò accade in quanto è reale, ormai, solo ciò che viene ad apparire. Più precisamente: ciò che si fa apparenza. Ecco perché l'espressione «realtà virtuale» non stupisce più: giacché il presunto ossimoro esprime l'esperienza concreta di un'indifferenza fra questi due termini.

2.2 Ciò ha conseguenze decisive anche riguardo al modo in cui oggi pensiamo lo *spazio* e il *tempo*. O meglio: riguardo al modo in cui queste stesse nozioni si trovano ad essere radicalmente risemantizzate, fino ad avere solo un labile legame con ciò che in passato esprimevano. Si parla infatti, oggi, di «cyberspazio» e di «tempo reale». Spazio e tempo, nella rete, non vengono affatto aboliti, ma trasformati. Il che significa che la rete ha il suo spazio, apre spazio, è, anzi, essa stessa spazio. E, analogamente, ciò che accade nella rete si svolge in una specifica dimensione temporale.

Ma che cosa vuol dire tutto questo, concretamente? Lo spazio della rete, in primo luogo, è uno spazio virtuale, in quel senso di 'virtuale' che abbiamo precisato poco sopra; è articolato secondo i rimandi dei link; è disposto nella forma del collegamento, cioè della connessione fra i vari nodi della rete. Esso è costituito, in generale, da punti ben definiti, i nodi, e da rimandi, che consentono l'apertura dei vari siti. È insomma uno spazio composto di altri spazi, a cui esso infinitamente rimanda, e insieme uno spazio che si apre infinitamente al suo interno, si approfondisce in se stesso, secondo la logica dell'ipertesto.

Tempo della rete, invece, è il tempo reale. Il che significa: il tempo dell'immediatezza, l'istante in cui sono racchiusi i vari percorsi possibili, che poi possono essere seguiti passando da un sito all'altro. La velocità di connessione fra questi testi compossibili, garantita da processori sempre più sofisticati, consente progressivamente di realizzare ciò che era già insito nella struttura dell'ipertesto: la simultaneità dei collegamenti. Da questo punto di vista, allora, i percorsi diacronici tendono asintoticamente alla sincronia. E la rete nel suo complesso ha la funzione di custodire in sé tutti questi processi tendenzialmente simultanei. Certo: si tratta di una simultaneità affatto potenziale, ma che risulta la condizione di possibilità di ogni connessione diacronica. Ecco perché – così come accade nella teoria della relatività e, pur in un senso diverso, nella riflessione dell'ultimo Heidegger – anche in Internet spazio e tempo rimandano strutturalmente l'uno all'altro.

D'altronde queste due nozioni non si riferiscono solamente a fenomeni che sono propri della rete e che risultano interni ai suoi processi. Esse possono infatti indicare anche tutto ciò che riguarda la relazione tra il programmatore, l'utente, il navigatore e la rete stessa. Vi sono dunque lo spazio e il tempo di Internet, quelli sperimentati da chi si muove nella dimensione virtuale della rete, e lo spazio e il tempo dell'uomo 'nel suo rapporto con' Internet. Quest'ultimo, nel quotidiano, vive per lo più in uno spazio familiare e pensa in successione, diacronicamente. Oggi, dunque, una riflessione sullo spazio e sul tempo non può esimersi dal

considerare, per un verso, l'intreccio fra il luogo dell'abitare quotidiano e l'ambito della connessione dei vari nodi, in un processo che può tendenzialmente aprirsi all'infinito, e, per altro verso, l'incontro fra il tempo diacronico dei vissuti umani e quel «tempo reale» che si caratterizza per la simultaneità dei processi multimediali.

2.3 Ma non sempre è facile distinguere umano e cibernetico, reale e virtuale, «interno» ed «esterno»⁶. Il perché lo abbiamo detto: il contenuto si realizza nel suo manifestarsi; il reale finisce per dissolversi nell'apparenza. Vengono a cadere dunque, nell'utilizzo della rete, alcune distinzioni tradizionalmente assunte. Ciò che viene meno, in definitiva, è la differenza tra naturale e artificiale. E se, da un lato, a un tale processo si accompagna una sensazione di potenza, dall'altro esso lascia un'impressione di sconcerto, suscita spaesamento. Ciò che sconcerta, in particolare, è il potere di assorbimento e di espansione di cui la rete e, più in generale, la dimensione del virtuale danno prova.

Un esempio indicativo in tal senso è dato dal modo in cui cambia, nell'età di Internet, il nostro rapporto con il corpo. Si verifica infatti un processo duplice: si assiste, da una parte, a una sorta di «virtualizzazione del corpo», alla quale si accompagna, dall'altra parte, una sorta di «incorporazione del virtuale»⁷. Sotto il primo titolo possono essere ricondotti tutti i fenomeni di simulazione che è possibile porre in opera nella «realtà virtuale»: dove ora quest'espressione va intesa, in senso stretto, come l'ambito nel quale ci si può inserire interattivamente con l'aiuto di particolari protesi ottico-tattili-auditive, all'interno di un ambiente tridimensionale generato dal computer. Il secondo caso riguarda invece situazioni come quelle legate al potenziamento e alla trasformazione del corpo anche e soprattutto mediante l'uso di protesi artificiali di vario tipo: non necessariamente attraverso trapianti e impianti, ma anche semplicemente sviluppando una dipendenza da certi strumenti di comunicazione, computer compreso.

Vengono così messe in questione, appunto, le tradizionali differenze tra «interno» ed «esterno», e tra «naturale» ed «artificiale». Ciò a cui si

⁶ Sembrano quasi acquisire nuovo senso, applicate a questo contesto, le parole usate dal Cusano (*De visione Dei*, 11) per esprimere il rapporto tra uomo e Dio: «Quando ti trovo come *virtus* che parimenti complica ed esplica, entro ad un tempo ed esco [...] Entro ed esco insieme, quando vedo che uscire è insieme entrare ed entrare è uscire».

⁷ Cfr. in proposito LÉVY, *Il virtuale* cit., pp. 17-23.

mira, attraverso le tecnologie multimediali, è in definitiva una progressiva 'integrazione': non più tra macchina e macchina; non più tra vari mezzi di comunicazione, ma tra macchina e uomo, tra artificio e natura, tra estraneo e familiare, tra proprio e altrui.

2.4 Infine, parallelamente a ciò, mutano anche i modi, le forme mentali, le prospettive in cui l'uomo si rapporta al mondo. All'approccio di tipo sequenziale che ad esempio la lettura richiede, e che abbisogna per il suo svolgimento della diacronia temporale, si sostituisce la simultaneità della rete e il procedere seguendo le infinite connessioni di essa. La 'connettibilità', insomma, viene posta al centro dei processi. E questo, lungi dall'essere un semplice dato di fatto, ben presto si trasforma in una condizione che deve essere perseguita e salvaguardata.

Muta dunque il senso stesso dell'essere in relazione. S'impone l'idea di una relazione aperta, sempre suscettibile di estendersi a nuovi collegamenti. Si tratta di una relazione a rete, in cui tutti i punti possono essere collegati a tutti gli altri. E dunque la localizzazione di un punto non è determinata all'occupazione di uno spazio fisico, ma dalle modalità delle sue connessioni possibili. Ecco perché, tenendo conto di questo, diventa sempre più decisivo – non solo in una chiave economica, ma soprattutto da un punto di vista etico – la questione dell'accesso alle connessioni⁸. È a partire da qui che, nel mondo di Internet, emerge il problema del *digital divide*, di cui fra poco ci occuperemo.

Prima, tuttavia, dobbiamo mettere in luce un ultimo aspetto relativo a come Internet cambia i nostri modi di pensare e di agire. Esso riguarda la categoria stessa dell'individuo. Per un verso, infatti, Internet, come rete delle connessioni possibili, sembra offrire sempre nuove possibilità di relazione, mettendo in crisi una caratterizzazione autonoma e conclusa dell'individuo. Per altro verso, però, l'individuo, nella sua relazione con la rete, viene concepito bensì come connettibile, ma pur sempre a partire da una sorta di isolamento di base che lo caratterizzerebbe preliminarmente. E da tale isolamento, che è il suo modo specifico di essere davanti al computer, egli può comunque partire per le sue avventure virtuali: cambiando *nickname*, gestendo in contemporanea più identità, rapportandosi simultaneamente a infiniti nodi.

È vero, dunque, che la rete promuove la socialità. Ma si tratta di una

⁸ Si veda, su tali questioni, il libro di J. RIFKIN, *L'era dell'accesso*, trad. it. di P. Canton, Milano, Mondadori 2000.

socialità ben determinata: che presuppone individui fra loro isolati, capaci certamente di vivere molteplici rapporti, anche se non si sa con quanto coinvolgimento emotivo. Giacché Internet, pur aprendo con la sua interattività nuovi orizzonti d'interesse per l'uomo, è pur sempre una dimensione da cui si possono prendere senza fatica le distanze: semplicemente spegnendo il computer.

3 *Problemi economici, sociali e politici nella gestione della rete: il caso del digital divide*

La rete, insomma, dischiude un ambito eticamente ambiguo. Luogo di opportunità e di rischi, essa ci mette di fronte a scelte ben precise riguardanti il suo utilizzo e la sua fruizione. Delle conseguenze di queste scelte, certo, siamo responsabili, e ci riconosciamo tali nella misura in cui colleghiamo tali conseguenze al nostro agire individuale e concreto. Meno siamo disposti a riconoscere una siffatta responsabilità personale per quanto riguarda alcuni aspetti di carattere più latamente sociale.

È questo il caso del *digital divide*, del divario tecnologico, a cui già abbiamo accennato. Per motivi economici, sociali e politici vi è infatti, nella geografia di Internet, un notevole squilibrio a livello mondiale fra luoghi che sono connessi al network e luoghi che non lo sono. E ad esso corrisponde un analogo squilibrio nelle opportunità d'innovazione tecnologica, di sviluppo delle conoscenze e di crescita economica che all'utilizzo della rete sono correlate. Si tratta, per di più, di un divario che, per la logica del progresso tecnologico, è destinato ad allargarsi in misura crescente, aprendo un vero e proprio abisso tra zone del mondo che sono collegate e zone che non lo sono, tra aree del pianeta adeguatamente attraversate dall'infrastruttura delle telecomunicazioni di Internet e aree che, se vogliono essere messe in rete, debbono usare i nodi altrui⁹.

Questo squilibrio, così come altri su scala mondiale, pone di fronte – allo stesso modo di altre conseguenze relative all'uso di Internet – a specifiche questioni morali. Che certo sono diverse a seconda del grado di

⁹ Si vedano i grafici, esplicativi anche se un po' datati, contenuti nel capitolo 8 del libro di M. CASTELLS, *Galassia Internet* (2001), trad. it. di S. Viviani, Milano, Feltrinelli 2002, e le considerazioni di A. MATTELART nel suo *Storia della società dell'informazione*, trad. di S. Arecco, Torino, Einaudi 2002 (2001).

coinvolgimento del soggetto agente. E dunque si può identificare, rispetto ad esempio al *digital divide*, una responsabilità diretta per la situazione vigente, quale quella che è propria di alcuni governi e delle grandi compagnie di telecomunicazione, e una responsabilità indiretta, anche labile, come quella di cui si può far carico il semplice utilizzatore della rete. Ma anche quest'ultima forma di responsabilità non va affatto trascurata: nella misura in cui in un'epoca, come quella contemporanea, caratterizzata da un'interconnettività sempre più forte – ed è questo, come ben sappiamo, uno dei caratteri distintivi della globalizzazione, comunque essa venga considerata – ogni soggetto connesso al *network* ha se non altro il potere di decidere, per parte sua, se continuare a sostenere o meno questa struttura di collegamenti. E lo può fare, oppure no, semplicemente collegandosi a sua volta, e in tal modo alimentando la possibilità di sempre nuove connessioni, o scegliendo, ad esempio, un determinato portale a discapito di altri. Di questi atti, e delle loro conseguenze, anche il mero fruitore della rete porta dunque la sua parte di responsabilità¹⁰.

4 Internet e lo spazio delle scelte

Internet, dunque, non è un destino: è una dimensione con la quale è possibile interagire. Solo per questo, perché nei suoi confronti vi è spazio per scelte e per assunzioni di responsabilità, esso può risultare oggetto di un'indagine di tipo morale. Più ancora, però, *Internet stesso* risulta, per la sua struttura, un luogo d'interazione. Esso in altre parole è l'ambito in cui ciascuno di noi, pur all'interno di certe condizioni, si trova a esercitare la propria libertà di scelta. Solo perciò, come vedremo, ha senso parlare anche di un'etica *in* Internet.

La struttura di Internet, ben lo si sa, è quella dell'ipertesto. E dunque, per muoversi nella rete, è necessario che ad ogni passo vengano prese delle decisioni. Infatti un ipertesto – per usare le parole di colui che ha coniato il termine: Theodor Nelson – è «un testo che si dirama e che consente al lettore di scegliere: qualcosa che si realizza al meglio davanti a uno schermo interattivo»¹¹. Ecco allora che davvero la rete, almeno agli

¹⁰ Ho approfondito questo aspetto nel mio saggio *Responsabilità nell'etica della comunicazione*, in «Teoria» 2004/1, fascicolo dedicato al tema «Responsabilità nelle etiche applicate».

¹¹ TH. NELSON, *Literary Machines* 90.1, Muzzio, Padova 1992, p. 2. Continua Nelson:

occhi del semplice fruitore, si presenta come una ramificazione di possibilità, come l'apertura possibile di sempre nuove connessioni. In effetti, è la capacità di connessione – la connettività, o anche, potremmo dire: la «connettibilità» – ciò che fa sì che Internet si configuri come uno spazio di scelte. Ed è a partire da qui che può concretarsi il potenziale di liberazione insito nella rete, consentendo a chi nella propria situazione non si può «muovere» liberamente – in un senso metaforico o letterale – di farlo almeno in uno spazio virtuale. Si comprende in tal modo l'intimo legame che sussiste, su di un piano politico, fra l'accesso a Internet e l'esercizio della democrazia.

Ma c'è di più. La struttura stessa della rete, in quanto essa è data dal collegamento di elementi, di «nodi» autonomi e separati, è tale da risultare aperta a sempre nuove integrazioni. Si tratta davvero di un *textum*, di un «tessuto» in costante espansione, sia da un punto di vista quantitativo che da una prospettiva qualitativa. Potremmo anzi dire che, da questo punto di vista, la rete si attua propriamente come un'«opera aperta».

A buon diritto, dunque, Internet si configura come luogo di decisioni: decisioni che, naturalmente, possono essere arbitrarie oppure orientate in base a principi o ad esigenze preventivamente assunti da chi è chiamato a prenderle. La rete, più ancora, diventa ambito per l'esercizio della propria libertà: di una libertà, certo, che deve fare i conti con condizioni ben precise. Dobbiamo allora chiederci quali sono tali condizioni. E vedere il modo in cui esse consentono alla libertà di scelta, richiesta dalla struttura ipertestuale, di realizzarsi concretamente.

Triplice, almeno a un primo sguardo, è la forma che assumono tali condizioni. Anzitutto appare chiaro che le scelte compiute all'interno dell'ambito di connettibilità risultano pur sempre inserite in una struttura rigida, ben definita, che è quella progettata dal programmatore. E con questa struttura, considerata di per sé, è possibile interagire solo parzialmente, almeno se si resta semplici navigatori. Ben diversamente accade se si hanno gli strumenti, tecnici e culturali, per programmare. In questi casi si possono compiere anche azioni di disturbo e di controinformazione, variamente motivate: come quelle compiute ad esempio in nome dell'etica hacker¹².

Spesso, poi, tali comportamenti sono attuati in base a un'ideologia

«Comunemente inteso un ipertesto è una serie di brani di testo tra cui sono definiti i legami che consentono al lettore differenti cammini».

¹² Cfr. P. HIMANEN, *L'etica hacker*, Milano, Feltrinelli 2001.

per la quale Internet è visto come luogo di liberazione soprattutto in virtù del facile accesso, che esso consente, a conoscenze e informazioni. Tutto questo è certamente vero, sebbene oggi cominciamo ad accorgerci che troppe informazioni non selezionate hanno lo stesso effetto di una sistematica disinformazione. Riflettendo più a fondo, però, va detto che proprio questa liberazione che Internet promette è una liberazione che si attua specialmente in un ambito virtuale. Anzi: proprio per questo motivo essa stessa sembra configurarsi, di fatto, come una liberazione virtuale. Per parlare allora di una liberazione vera, reale, promossa da Internet, bisogna considerare, più che quanto accade nella rete, ciò che la rete consente di modificare nella sfera dei rapporti economici e sociali. E a questo proposito non è certo il caso d'insistere sull'incidenza delle nuove tecnologie, ad esempio, nelle trasformazioni del lavoro. Ma con ciò si ripropongono anche quegli squilibri di cui abbiamo parlato, precedentemente, in relazione al *digital divide*.

Infine bisogna rilevare che le scelte che ci vengono proposte all'interno della rete, in virtù del suo carattere di ipertesto, si collocano su di un piano dominato, potremmo dire, da criteri di semplice spiegazione. Si naviga in Internet, si utilizza questo strumento, lo si modifica, per quanto possibile, a seconda delle nostre esigenze. Ma non ci si chiede, nel far ciò, 'qual è il senso' delle diverse procedure fra le quali esercitiamo la nostra capacità di scelta. Anche perché, se lo facessimo, ci accorgeremmo ben presto che un tale senso, di per sé, la rete non è in grado di fornirlo. Esso va cercato altrove, e sulla sua base vanno motivate le proprie decisioni all'interno dell'ipertesto. Oppure, quando siamo stanchi di muoverci in Internet, possiamo compiere un gesto altrettanto insensato: il gesto di disconnetterci, o addirittura, come si è detto, di spegnere il computer.

Se dunque le cose stanno così, se vi è una serie di condizioni complesse con le quali interagiamo e che dobbiamo accogliere quando navighiamo in Internet, dobbiamo allora chiederci, in definitiva, qual è l'idea di libertà che, propriamente, emerge in tutti questi processi. E a questa domanda è necessario rispondere, se vogliamo che sia i nostri comportamenti *nei confronti* della rete, sia quelli *all'interno* della rete stessa possano essere pensati in una chiave etica. A ben vedere, infatti, la libertà che qui è in gioco è intesa, comunemente, soprattutto come libertà di fruizione: sulla linea di quell'ampliamento delle possibilità e di quel conseguente aumento di potere che l'utilizzo di Internet rende possibile. Quest'idea, d'altronde, è in linea con la trasformazione del concetto stesso di 'libertà' che lo sviluppo della tecnica rende possibile. E si collega altresì all'esigenza – tipica della logica del consumo – d'intendere la libertà di

scelta come ampliamento dell'offerta proposta nel quadro di un menu ideale: un ampliamento che, come abbiamo visto, la rete sembra in grado di poter estendere all'infinito.

Questa, però, non è l'unica idea di libertà che si può sperimentare, neppure all'interno della rete. Bisogna allora fare i conti fino in fondo con questa concezione, comprendendone la parzialità, e rendendoci conto che altri sensi di 'libertà' e di 'scelta' possono essere elaborati, proprio in relazione a ciò che Internet rende possibile. Fra questi vi sono, in primo luogo, i significati propri di un agire inteso non già come qualcosa che, unicamente, risulta in costante espansione, bensì come ciò che è in grado, liberamente appunto, anche di (auto)limitarsi. Ma questa, di nuovo, altro non è se non l'idea di 'libertà responsabile'.

5 Etica in Internet

Con queste analisi relative, da un lato, a Internet come spazio delle scelte e, dall'altro, alle condizioni per esercitare quella libertà che la rete consente, siamo già passati, di fatto, dall'«etica di Internet» all'«etica in Internet»: cioè a quell'insieme di comportamenti che bisogna adottare quando si utilizzano le possibilità del Web e quando si naviga in rete. A questo proposito sono state già elaborate puntuali indicazioni e specifiche prescrizioni, che si fondano su argomenti di carattere giuridico, su motivi di opportunità, su sollecitazioni morali oppure su istanze collegate al semplice galateo. C'è stato addirittura chi, come Arlene Rinaldi, ha tentato di fissare tali prescrizioni, elaborando «dieci comandamenti» a cui si deve rifare chiunque utilizzi un computer¹³.

A ben vedere, però, in questo come in altri casi analoghi, più che di «etica» si deve parlare di «etichetta». O più precisamente, volendo usare un neologismo coniato per la bisogna, di *netiquette*. Ben diverso,

¹³ Cfr. <<http://www.fau.edu/rinaldi/net/intro.html>>. I comandamenti sono: 1. Non userai un computer per danneggiare altre persone; 2. Non interferirai con il lavoro al computer di altre persone; 3. Non curioserai nei file di altre persone; 4. Non userai un computer per rubare; 5. Non userai un computer per portare falsa testimonianza; 6. Non userai o copierai software che non hai dovutamente pagato; 7. Non userai le risorse di altri senza autorizzazione; 8. Non ti approprierai del risultato del lavoro intellettuale altrui; 9. Pensarai alle conseguenze sociali dei programmi che scrivi; 10. Userai il computer in modo da dimostrare considerazione e rispetto.

invece, è il caso di quei codici di autoregolamentazione che, anche nel caso della rete, cominciano da più parti ad essere elaborati coinvolgendo gli operatori del settore e le aziende erogatrici dei servizi. Tali codici riguardano, per ora, alcuni degli abusi più di frequente perpetrati nel Web, oppure cercano di salvaguardare specifiche categorie di potenziali utenti: ad esempio i minori.

Sempre dal versante della tutela degli utenti da qualche anno si è cominciato a tener conto, sempre di più, delle questioni di accessibilità e di usabilità della rete, in modo da poter effettivamente estendere le sue potenzialità ad un numero sempre più ampio di fruitori. Ciò in particolare è avvenuto quando si è riconosciuto in Internet uno strumento fondamentale per la gestione della Pubblica Amministrazione. Ma bisogna sempre aver chiaro che questa riconversione in rete delle procedure amministrative, così come la correlata esigenza di garantire l'accesso a Internet al maggior numero di cittadini, possono adeguatamente realizzarsi solo se sono chiare le motivazioni da cui tali iniziative sono animate. E queste motivazioni sono non già di carattere semplicemente tecnico, bensì, soprattutto, di tipo morale.

Accanto alla *netiquette* o galateo in Internet, accanto ai codici che mirano alla salvaguardia o al coinvolgimento di particolari categorie di utenti, vi sono – e sono ben rappresentati in rete – regolamenti che intendono dare indicazioni sui modi in cui forme specifiche di testi multimediali debbono essere redatte. In tali regolamenti, cioè, vengono individuati quei criteri che – nel caso ad esempio di un *blog* o di una *mail*, del collegamento a una *chat* o, più in generale, a un determinato sito – consentono un'elaborazione corretta e soprattutto rispettosa degli altri utenti. Si tratta di criteri che debbono essere adottati se si vuole continuare a far parte di una comunità: altrimenti si può rischiare di esserne esclusi.

D'altra parte, una tale soluzione – che richiede la presenza di un moderatore in grado di sanzionare le trasgressioni – non può essere estesa a tutto il Web. Infatti, dal momento che Internet non è gestito da un ente supervisore, unico e riconosciuto, non è possibile per il momento far riferimento a un'istanza generale di controllo. Forse, d'altronde, è anche bene così, perché questa istanza suprema potrebbe facilmente trasformarsi nel Grande Fratello di orwelliana memoria.

Nell'attuale situazione, perciò, ogni utente finisce per essere in qualche modo chiamato ad assumersi le proprie responsabilità e, soprattutto, a trovare in se stesso le motivazioni che consentono di seguire un comportamento corretto. Vi sono, certamente, siti a cui possono essere segnalati gli eventuali abusi. Vi è, sul piano penale, una legislazione che punisce certe pratiche commesse con l'utilizzo della rete. Ma nella

misura in cui Internet non può essere delimitato all'interno di precisi confini geografici, è difficile applicare nel suo caso leggi valide entro la giurisdizione degli Stati nazionali. Per il futuro, più di quanto non si sia fatto sinora, dovranno dunque essere stabiliti accordi internazionali per la regolamentazione della rete e del suo corretto utilizzo.

Ma allo stato attuale, nella misura in cui è oltremodo difficile, per i motivi accennati, comminare sanzioni effettivamente applicabili, siamo ancora demandati a meccanismi di autoregolamentazione del sistema e, appunto, al senso di responsabilità delle aziende, dei programmatori, degli utenti. Ciò vale, lo ripeto, in casi e situazioni anche sensibilmente diversi fra loro. Giacché un conto sono violazioni perseguibili anche penalmente: si pensi all'uso di siti web da parte di pedofili, speculatori o truffatori; un conto sono le violazioni del galateo della rete: ad esempio l'inondare l'altrui casella di posta elettronica di innumerevoli messaggi non richiesti (il cosiddetto *spamming*), oppure le questioni che si collegano ai modi, non sempre trasparenti, di interagire all'interno di una *chat*; un conto, infine, sono i problemi di carattere più propriamente morale. E per una corretta definizione di essi, per il loro adeguato approfondimento, molto ancora dev'essere fatto.

6 Conclusione

Da quanto abbiamo fin qui esposto, in ogni caso, risulta che, per regolamentare i comportamenti all'interno della rete, non bastano i codici o le sanzioni da essi previste. Bisogna invece fornire le motivazioni, indicare il perché è giusto compiere determinati atti piuttosto che altri. E questo è ciò che l'etica *in* Internet è chiamata a fare, con il suo riferimento alla responsabilità di ciascun soggetto collegato alla rete.

Anche questo passaggio, però, non è sufficiente. Per giustificare l'adozione di determinati comportamenti l'etica *in* Internet deve a sua volta essere collocata nel contesto più generale delle problematiche della rete e dei nodi etici che esse coinvolgono – l'etica *di* Internet – e, a partire da qui, debbono essere esibiti quei principi generali ai quali di volta in volta ci si richiama nelle scelte concrete. E fondare tali principî, in maniera adeguata, è compito dell'etica generale¹⁴.

¹⁴ Per un inquadramento complessivo dei problemi generali affrontati dall'etica della comunicazione, nel suo rapporto con l'etica generale, mi permetto di rinviare al mio libro *Etica della comunicazione*, Roma, Carocci 2006.

Insieme a tutto ciò, e come condizione del suo realizzarsi, va tuttavia fatta emergere, ed effettivamente sperimentata, una consapevolezza sia dei limiti dell'agire umano, sia del carattere altrettanto limitato di quegli stessi prodotti della tecnica che, pur mirando a rendersi autonomi e ad autoalimentarsi, dall'uomo ancora dipendono. È in questo senso del limite che, a ben vedere, risiede la radice ultima della responsabilità. E dunque di un'etica capace di gestire quella libertà che ne è il fondamento. Anche di fronte a fenomeni che, al pari di Internet, sono in grado di aprire nuove possibilità e orizzonti inediti per il mondo in cui viviamo.

ADRIANO FABRIS

'Archiviare' il Web: iniziative e progetti

1 Il Web come fonte storica

1.1 Il Web è un archivio?

Da quando, circa un decennio fa, sono state messe a punto le tecnologie su cui si fonda, il World Wide Web è diventato un formidabile veicolo di diffusione di informazione e di prodotti culturali e d'intrattenimento, composti da testi, foto, filmati, brani musicali e da una congerie difficilmente circoscrivibile di altri materiali digitali. All'immaginario collettivo la Rete appare come un enorme e sconfinato archivio, al quale attingere per ogni genere di necessità o curiosità informativa. La metafora di Internet come immenso deposito della memoria e della cultura dell'umanità, insomma come una sorta di sterminato archivio globale, è ormai diventata da tempo uno dei luoghi comuni più diffusi ed abusati della contemporaneità.

Ma la metafora dell'archivio è davvero in grado di rappresentare efficacemente la natura del Web? A guardar bene le cose non stanno propriamente così. In realtà, dell' 'archivio' il Web ha dimostrato, fino adesso, di avere ben poche caratteristiche. Soprattutto gliene manca una essenziale: la ragionevole aspettativa di ritrovarvi, dopo un certo lasso

Nelle more della pubblicazione nuovi progetti di archiviazione del Web sono stati avviati, così come sono cresciuti i contributi sul tema. Né degli uni né degli altri si potuto tener ovviamente conto. Per quanto concerne la bibliografia corre tuttavia l'obbligo di ricordare almeno i seguenti volumi: A. BROWN, *Archiving Websites: A Practical Guide for Information Management Professionals*, London, Facet 2006; *Web Archiving*, ed. by J. Masanès, Berlin-New York, Springer 2006.

Il presente contributo costituisce una rielaborazione di un mio precedente saggio dal titolo *Una memoria fragile: il Web e la sua conservazione*, in *La storiografia digitale*, a cura di D. Ragazzini, Torino, Utet Libreria 2004, pp. 100-127. I siti web segnalati erano accessibili alla data del 28 febbraio 2006.

di tempo, ciò che ad esso è stato affidato. Un archivio, in fondo, serve principalmente a garantire nel tempo una trasmissione, possibilmente ordinata, dei documenti che lo costituiscono e delle informazioni di cui questi sono portatori. Il Web sembra avere avuto fino adesso una funzione esattamente opposta: quella di rendere disponibile informazione il più possibile 'fresca' ed aggiornata a ritmi sempre più contratti che tendono all'istantaneità o, come si suol dire, al tempo reale. Poco invece ci si è fino adesso curati della sua stabilità e permanenza nel tempo e del destino di quanto, ritenuto ormai superato, viene eliminato o sostituito.

I siti web e le singole pagine di cui questi si compongono sono intrinsecamente instabili e fondamentalmente volatili. Le stime relative alla durata media di siti e pagine web sono tutt'altro che concordi, probabilmente non molto affidabili, ma certamente indicative di un fenomeno e di una tendenza. Quelle della fine degli anni '90 calcolavano in 44 giorni il tempo medio di vita di una pagina web, mentre stimavano al 44% la percentuale dei siti che non era accessibile da un anno all'altro¹. In altre stime, la durata media delle pagine web è stata valutata in sessanta oppure, più recentemente, in cento giorni². Ma al di là dell'esattezza statistica di queste cifre, è l'esperienza comune e quotidiana del navigatore di Internet che conferma quanto sia tutt'altro che raro scoprire che un determinato sito, già visitato in precedenza oppure 'linkato' da altri siti o segnalato altrove, non esiste più. In compenso, al medesimo navigatore capita sempre

¹ P. LYMAN, *Archiving the World Wide Web*, in *Building a National Strategy for Digital Preservation*, Council on Library and Information Resources and Library of Congress, aprile 2002, <<http://www.clir.org/pubs/reports/pub106/web.html>>.

² Si tratta di stime ricorrenti, formulate in più occasioni da vari specialisti del settore, come B. KAHLE dell'Internet Archive (sul quale vedi sotto) e riprese sulla stampa in molteplici occasioni, cfr., ad esempio, per la prima stima *British Library Attempts to Archive the Net*, in «Library Journal», 4 marzo 2002, <<http://libraryjournal.reviewsnews.com/index.asp?layout=article&articleid=CA206461&publication=libraryjournal>>; per la seconda, cfr. R. WEISS, *On the Web, Research Work Proves Ephemeral. Electronic Archivists Are Playing Catch-Up in Trying to Keep Documents From Landing in History's Dustbin*, November 24, 2003; page A08, anche online, <<http://www.washingtonpost.com/ac2/wp-dyn/A8730-2003Nov23?language=printer>>. Per un'analisi sulla stabilità dei siti web attraverso il monitoraggio nel corso del tempo di un campione casuale cfr. W. KOEHLER, *A Longitudinal Study of Web Pages Continued: a Consideration of Document Persistence: a Report after Six Years*, in «Information Research», IX, 2 gennaio 2004 (<<http://InformationR.net/ir/9-2/paper174.html>>).

più spesso di imbattersi in avvisi che indicano l'inesistenza o la scomparsa della pagina cercata – cioè l'inquietante «error 404: page not found» dell'http, l'Hypertext Transfer Protocol – col ricorso a soluzioni grafiche e a motti ammiccanti od ironici, quasi a voler attenuare la delusione o prevenire l'irritazione per la sgradevole scoperta e, nello stesso tempo, ammonire che per navigare nella rete bisogna armarsi di una buona dose di pazienza e di rassegnazione³. C'è chi, come il Marxist Internet Archive, prova a filosofeggiare sulla pagina mancante («the dialectics of opposites: you found much but this, you did not»)⁴; chi invita invece a prendersi una boccata d'aria fresca e non stare troppe ore attaccato allo schermo del computer («The page cannot be displayed because you need some fresh air. [...] Your ISP has told us you're on-line several hours every day, and that just ain't right»)⁵ oppure chi utilizza il messaggio d'errore per fare satira sulla spedizione americana in Iraq («These Weapons of Mass Destruction cannot be displayed The weapons you are looking for are currently unavailable. The country might be experiencing technical difficulties, or you may need to adjust your weapons inspectors mandate»)⁶.

Ma i siti web non hanno solo un'esistenza effimera, sono anche, per loro natura, estremamente dinamici, subiscono cioè trasformazioni, talvolta molto rapide, nella grafica, nelle modalità di presentazione dei materiali pubblicati e, soprattutto, nel contenuto. I mutamenti talvolta sono espliciti e ben visibili, talaltra quasi nascosti e difficili da cogliere, ma non per questo meno rilevanti, soprattutto quando alterano il significato sostanziale dei messaggi veicolati. Michael Gorman, uno dei più autorevoli rappresentanti delle scienze bibliografiche contemporanee, ha acutamente notato che i siti web con contenuti continuamente mutevoli [...] non hanno alcun parallelo nel mondo della stampa o dell'editoria di altri oggetti tangibili. Assomigliano

³ Per una storia ed una illustrazione delle varie implicazioni dell'error 404 cfr. il sito *404 Research Lab*, <<http://www.plinko.net/404/default.asp>>.

⁴ «Marxists Internet Archive: File Not Found», <<http://www.marxists.org/admin/errors/404.htm>>.

⁵ Cfr. la pagina costruita sulla falsariga del messaggio di errore di Internet Explorer per l'«error 404: page not found» nel blog canadese 'Mamselle.ca', <<http://www.mamselle.ca/error.html>>.

⁶ Cfr. <<http://www.coxar.pwp.blueyonder.co.uk/>>. Questa e l'altra pagina citata nella nota precedente sono segnalate in WEISS, *On the Web, Research Work Proves Ephemeral* cit.

di più a quelle curiose strutture che si vedono per le vie di Parigi – erette per ricoprirle di manifesti. I manifesti cambiano, si strappano e vengono ricoperti da nuovi manifesti, vengono staccati e deturpati con graffiti, tanto che il contenuto e l'effetto visivo sono diversi di settimana in settimana. Solo il sito resiste.

Proprio questi caratteri del Web – aggiunge Gorman – mettono radicalmente in discussione la convenienza di prospettarne la conservazione e la catalogazione:

finché il contenuto è così instabile e la forma volubile e finché i siti stessi, diversamente dalle strutture parigine possono svanire da un giorno all'altro, è molto difficile, se non impossibile conservarli e sottoporli al controllo bibliografico. Mi sorge una domanda forse eretica: Ma, è importante? Il contenuto di quei siti ha qualche valore durevole? È certo che gli autori di quel contenuto gli attribuiscono pochissimo valore, dato che essi stessi non fanno alcun tentativo per conservarlo⁷.

Non si può negare che interrogativi del genere siano legittimi e non privi di un qualche fondamento. Eppure, nonostante l'instabilità e la volatilità del Web, anzi, paradossalmente proprio per questi caratteri sta da qualche tempo crescendo la consapevolezza della necessità di elaborare e mettere in campo strategie che permettano di conservare il Web, di 'archiviarlo' in qualche modo, proprio perché diventa sempre più evidente che l'avvento di Internet ha costituito uno dei fenomeni più rilevanti affermatosi negli anni Novanta del secolo passato e che, da allora, il Web svolge un ruolo fondamentale nella società, nell'economia, nella cultura del nuovo millennio⁸.

1. 2 *Web e società contemporanea*

In tutti questi anni il Web non è stato soltanto uno strumento di aggregazione e di comunicazione di una messe sterminata di documenti

⁷ M. GORMAN, *Le risorse elettroniche. Quali vale la pena conservare e qual è il loro ruolo nelle raccolte della biblioteca? Le risorse elettroniche. Definizione, selezione e catalogazione*. Atti del convegno internazionale (Roma, 26-28 novembre 2001), a cura di M. Guerrini, con la collaborazione di S. Gambari e L. Sardo, Milano, Editrice Bibliografica 2002, p. 95.

⁸ Per una introduzione generale alle problematiche relative al *Web archiving* cfr. LYMAN, *Archiving the World Wide Web* cit.; M. DAY, *Collecting and Preserving the World Wide Web. A feasibility study undertaken for JISC and the Wellcome Trust*, Versione 1.0, 25 febbraio 2003, <http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf>.

e informazioni, generati direttamente in formato digitale oppure migrati su quel formato attraverso processi di codifica digitale di materiali originariamente esistenti su supporti tradizionali. Esso è stato anche il terreno di affermazione di specifiche forme di cultura che hanno messo a punto codici, linguaggi e modalità di espressione originali che non hanno equivalenti nel mondo degli altri media. Basti pensare, tanto per fare un esempio, ai Weblog, o blog, fenomeno in corso di rigoglioso sviluppo, di dimensioni enormi, seppure difficili da determinare con precisione, come tutto ciò che riguarda Internet, tanto che si calcola attorno ai due milioni il numero di quelli che sono costantemente aggiornati⁹. I blog costituiscono uno strumento di difesa della libertà di espressione individuale, secondo uno stile che in Internet ha trovato sempre ampio spazio. Ma, come punti di aggregazione e di snodo della comunicazione di rete, essi consentono di contenere in modo intelligente quel fenomeno, tipico non solo di Internet ma dell'intera società contemporanea, che è l'*information overload*, offrendo un terreno del tutto nuovo ed originale di costruzione di un'opinione pubblica consapevolmente e democraticamente informata¹⁰. Ma essi rappresentano anche un cantiere dal quale emergono forme di creatività letteraria nuove ed originali¹¹. Come è stato notato i blog:

rappresentano un forma ben riconoscibile ed originale di comunicazione che sta avendo un impatto ben verificabile sulle vicende umane all'inizio del XXI secolo [...] [la conservazione di] un'ampia selezione di ogni tipo di blogs ben

⁹ R. ENTlich, *Blog Today, Gone Tomorrow? Preservation of Weblogs*, in «RLG DigiNews. RLG's online newsletter for digital imaging and preservation», VIII, 4, 15 agosto 2004, <http://www.rlg.org/en/page.php?Page_ID=19481>, calcolava in circa due milioni i blog aggiornati costantemente. Stima prudente a vedere le cifre indicate dal più noto motore di ricerca della blogosfera, Technorati, <<http://www.technorati.com/about/>>, che nel gennaio 2006 fa ascendere a più di ventisette milioni i blog indicizzati. In Italia i blog sarebbero, più o meno alla stessa data, attorno ai quattrocentomila, cfr i numeri di dicembre [2005], nel blog *Come si fa un blog*, di Sergio Maistrello, <<http://www.come-sifaunblog.it/categoria/numeri/>>.

¹⁰ Insiste sull'importanza dei blog quale strumento per elevare il tasso di democrazia effettiva delle società contemporanee, G. GRANIERI, *Blog Generation*, Roma-Bari, Laterza 2005.

¹¹ Sui blog come fenomeno letterario cfr. *La notte dei blogger*, a cura di L. Lipperini, Torino, Einaudi 2004.

scelti sarebbe indispensabile per fornire un'idea generale dei blog come fenomeno sociale e culturale e per aiutare lo storico del futuro a capire la nostra epoca. Purtroppo una raccolta mirata di blog così come della maggior parte dei materiali digitali effimeri, non sta ottenendo molta attenzione da parte di bibliotecari e archivisti¹².

Sarebbe tuttavia riduttivo attribuire la rilevanza che il Web ha assunto negli ultimi anni soltanto a fenomeni che, nati sulla rete, la hanno assunta come unico orizzonte, restando per lo più in essa circoscritti e intrecciando limitati rapporti con la realtà complessiva. Al contrario, nell'ultimo decennio si sono affermati in modo sempre più deciso usi di Internet, che tendono ad integrare la rete nella vita sociale di individui e collettività, riconfigurando comportamenti, stili di comunicazione, modalità di socializzazione, forme di organizzazione e di iniziativa civile e politica sulla base delle logiche e degli strumenti messi a disposizione dalla rete. Esempio è, fra tutte le esperienze che si potrebbero citare, quella dei nuovi movimenti che si sono sviluppati, in ambito sia internazionale che nazionale, fra lo scorcio del Novecento e l'inizio di questo. Senza i siti web e la rete in generale, non sarebbe possibile concepire e spiegare i caratteri e le dinamiche che distinguono movimenti come quello mondiale antiglobalizzazione oppure quelli italiani di protesta della società civile contro il governo Berlusconi, da quelli affermatasi nei decenni precedenti. Come è stato notato,

per questi movimenti Internet è più di uno strumento da utilizzare solo perché è a portata di mano. Essa si accorda perfettamente con i caratteri basilari dei tipici movimenti sociali che emergono nell'era dell'informazione. Ed è proprio perché essi hanno trovato il loro appropriato strumento di organizzazione che questi movimenti si sviluppano ed aprono nuove vie di trasformazione sociale, che, a loro volta, rafforzano Internet nel ruolo di loro privilegiato mezzo di comunicazione.¹³

Di fronte alla crisi storica di legittimazione e di capacità di mobilitazione dei tradizionali strumenti verticalmente integrati di mediazione politica e di formazione del consenso, quali i partiti e i sindacati, di fronte alla

¹² Cfr. ENTlich, *Blog Today, Gone Tomorrow? Preservation of Weblogs* cit. Sui Weblog cfr. anche il capitolo ad essi dedicato in M. CALVO-F. CIOTTI-G. RONCAGLIA-M. A. ZELA, *Internet 2004. Manuale per l'uso della rete*, Roma-Bari, Laterza 2003, pp. 269-296.

¹³ M. CASTELLS, *The Internet Galaxy. Reflections on the Internet, Business, and Society*, Oxford, Oxford University Press 2001, p. 139.

crisi dello stesso modello organizzativo centralizzato e verticistico, che tali strutture hanno tradizionalmente incarnato, Internet e World Wide Web sembrano insomma offrire nuovi canali di aggregazione e di iniziativa civile, che, basandosi sulla orizzontalità dei collegamenti e sull'«identità fra la comunicazione e l'organizzazione»¹⁴, risultano più rispondenti ad una realtà di movimento policentrica e reticolare¹⁵.

All'interno di contesti completamente diversi, i siti web divengono strumento – e recano, quindi, testimonianza – del tentativo di attivare nuove forme di rapporto fra le istituzioni e i cittadini, mettendo questi ultimi non solo «in grado di conoscere nel dettaglio le scelte politiche, gli strumenti e i risultati di monitoraggio e valutazione delle stesse, l'agenda di lavoro degli organi di governo, le discussioni, i documenti intermedi, i documenti di bilancio» ma anche di partecipare al processo di assunzione delle scelte politiche e amministrative o di influenzarle più o meno direttamente. Basti pensare allo sviluppo che le reti civiche, promosse da molte amministrazioni locali in Italia e in altri paesi europei ed extraeuropei, hanno avuto negli ultimi anni e al possibile rilievo che in futuro potranno avere i materiali che sui loro siti si sono sedimentati per valutare i caratteri effettivi e i risultati delle politiche di partecipazione che i loro promotori hanno cercato di incoraggiare per rispondere alla crisi della rappresentanza politica da cui le istituzioni tradizionali sono state investite negli ultimi decenni¹⁶. Inoltre, sempre più frequentemente uffici ed enti pubblici utilizzano i propri siti web come strumento di comunicazione con il pubblico, notificando ufficialmente attraverso di essi le proprie decisioni e portando a conoscenza

¹⁴ L. PICCI, *La sfera telematica*, Bologna, Baskerville 1999, p. 129.

¹⁵ Alle tematiche del rapporto fra movimenti e tecnologie telematiche e ai problemi di conservazione dei materiali digitali prodotti dai movimenti è stato dedicato alla fine del 2004 un convegno organizzato dalla Fondazione Giangiacomo Feltrinelli di Milano, dal titolo «Gli archivi del presente: dal documento tradizionale al documento digitale. Fonti per una storia dei movimenti sociali contemporanei: informazioni e testi delle relazioni sono reperibili sul sito della Fondazione, <<http://www.feltrinelli.it/Fondazione/convegni-interna-archivi-it.htm>>; cfr. anche S. MENAFRA, *Se la memoria corre sul file*, «Il manifesto», 5 dicembre 2004, anche *online*, <<http://www.ilmanifesto.it/g8/dopogennaio/41b488c002349.html>>.

¹⁶ Per una analisi critica delle esperienze italiane di rete civiche cfr. A. C. FRESCHI, *La società dei saperi. Reti virtuali e partecipazione sociale*, Roma, Carocci 2002, in particolare pp. 91-136 (per la citazione p. 127). Per uno sguardo sintetico a livello internazionale: CASTELS, *The Internet Galaxy* cit., pp. 144-158.

degli interessati bandi, scadenze e iniziative di ogni genere. Ma se ne servono anche per accogliere domande e istanze, informazioni e pareri, sollecitazioni e proteste. Internet si propone insomma come uno degli strumenti privilegiati di attuazione di progetti di *e-government*, che nelle intenzioni delle pubbliche amministrazioni della generalità dei paesi europei, dovrebbero avvicinare i cittadini alle istituzioni e metterle in grado di fornire risposte più rapide ed efficienti ai loro bisogni. Sui siti web, o grazie alla loro mediazione, si vengono così a sedimentare *corpus* di materiali documentari, che integrano o rimpiazzano quelli che solitamente si accumulano negli archivi delle istituzioni.

Non si tratta che di pochi, limitati esempi che mostrano bene come i materiali che circolano e si stratificano nella rete costituiscano rappresentazioni eloquenti di eventi e fenomeni della società contemporanea e che, come tali, essi siano destinati a diventare fonti di primaria importanza per lo storico del futuro – sempre che si riesca a conservarne e tramandarne alle future generazioni componenti significative¹⁷.

2 *Problemi concettuali, difficoltà pratiche*

2.1 *Quanto è grande il Web?*

Nonostante che la consapevolezza dell'opportunità di mettere in campo progetti di conservazione a lungo termine di siti e pagine web vada crescendo, i problemi concettuali e gli interrogativi di ordine pratico sono ben lungi dall'essere stati risolti in modo soddisfacente e secondo direzioni univoche. Al contrario, su molti nodi cruciali la discussione resta decisamente aperta. Se l'idea stessa di 'congelare' una realtà in perenne mutamento presenta di per se stessa non pochi elementi di contraddittorietà, difficoltà non indifferenti derivano dalla dimensione enormemente dilatata ed in continua e rapida espansione del Web, che pone interrogativi di non facile soluzione. Quanto sia grande il Web è molto difficile calcolarlo. Il Web Characterization Project dell'OCLC,

¹⁷ Per una prima informazione sui problemi della conservazione a lungo termine dei materiali digitali – sui quali non ci soffermiamo in questo contributo, ma che sono un nodo cruciale per rendere realmente possibile il ricorso ad archivi, documenti e ad altri 'oggetti' digitali nel lavoro storiografico – cfr. la bibliografia e la documentazione sul sito del progetto europeo ERPANET, *Electronic Resource Preservation and Access Network*, <<http://www.erpanet.org>>.

che cerca di stabilire quanti indirizzi IP – l'indirizzo numerico, composto di quattro gruppi di tre cifre ciascuno, che identifica i computer connessi alla rete –, ospitano *server* Web, ne ha individuati, nel 2002, circa 9 milioni. Rispetto al 1998, quando erano circa due milioni e seicentomila, essi sarebbero quindi quasi quadruplicati¹⁸. Una stima del genere quantifica, come si è detto, i *server* Web, ma non i siti, che possono essere molti di più, dato che lo stesso *server* può ospitarne molteplici. Secondo una metodologia di rilevazione, che tiene conto di quest'ultima possibilità, i siti web ascenderebbero, nel febbraio 2006, a circa 76 milioni¹⁹. Altre valutazioni prendono a base di calcolo le singole pagine web, dando esiti che appaiono, per le cifre che ne emergono, indubbiamente più impressionanti. Secondo uno studio della School of Information Management and System dell'Università di California a Berkeley il Web consisteva, alla fine del 2000, di 2 miliardi e cinquecento milioni di pagine, che crescevano ad un tasso medio di 7 milioni e 300 mila al giorno²⁰. D'altronde il numero delle pagine indicizzate che Google indicava fino a qualche tempo è giunto a superare, nel corso del 2005, gli 8 miliardi.

Va tenuto comunque conto che i *crawler* dei motori di ricerca che esplorano la rete, catturando i siti web, non possono raggiungere le pagine isolate, scarsamente o niente affatto 'linkate' ad altre, le quali, secondo alcuni calcoli, costituiscono almeno il 30% del Web. Inoltre, resta al di fuori della loro portata il cosiddetto *deep Web* o 'Web profondo', che secondo uno studio di qualche anno fa interessava circa 200 mila siti e aveva una dimensione fra le 400 e le 550 volte maggiore di quello 'di superficie' quantificabile in circa 550 miliardi di 'documenti'²¹, ma che adesso è

¹⁸ «The Web Charaterization Project» è consultabile all'URL <<http://www.oclc.org/research/projects/archive/wcp/>>.

¹⁹ Questo tipo di rilevazione è effettuata dalla società inglese di servizi Internet Netcraft: vedine i criteri e i risultati rispettivamente alle pagine 'Survey Mechanics', <<http://www.netcraft.com/Survey/mechanics.html>> e 'February 2006 Web Server Survey', <http://news.netcraft.com/archives/Web_server_survey.html>.

²⁰ Cfr. P. LYMAN-H. R. VARIAN, *How Much Information?*, ottobre 2000, <<http://www.sims.berkeley.edu/research/projects/how-much-info/>>; una versione più recente del medesimo studio non contiene valutazioni sul numero delle pagine web: cfr. *How Much Information?* 2003, ottobre 2003, <<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>>.

²¹ Cfr. M. K. BERGMAN, *The Deep Web: Surfacing Hidden Value*, in «Journal of Electron-

certamente molto più grande anche a causa delle tecniche prevalenti nella costruzione di siti web, che alla pubblicazione di pagine html statiche, prediligono la creazione *on the fly* di pagine dinamiche i cui elementi costitutivi sono memorizzati all'interno di apposite banche dati.

Di fronte a dimensioni di questo genere, che suggeriscono quanto il Web costituisca ormai una componente ineliminabile del più generale fenomeno di inflazione quantitativa dell'informazione che caratterizza la nostra epoca, ogni ipotesi di conservazione nel tempo dei materiali che lo compongono non può non confrontarsi con il problema dell'elaborazione di strategie di selezione in grado di individuare ciò che vale realmente la pena di tramandare nel tempo. Nonostante, infatti, le capacità di stoccaggio dei supporti elettronici ed ottici abbiano moltiplicato in modo esponenziale la quantità di informazione memorizzabile (e siano all'origine dell'*overload* informativo cui abbiamo appena ora accennato), la necessità di mettere in atto processi di selezione documentaria, non viene meno neppure adesso, quando, al contrario l'esplosione dell'informazione porta con sé più che mai ridondanze e ripetizioni, l'incremento del superfluo, povertà di contenuti piuttosto che ricchezza. E soprattutto rischia di annegare l'informazione effettivamente significativa in un 'rumore' di fondo che nasconde ed appiattisce ogni gerarchia di rilevanza. Da sempre la distruzione è stata il paradosso inevitabile di ogni strategia, più o meno consapevole, di conservazione. E ciò deriva non tanto o soltanto da ragioni di economia di spazio e di costi, quanto dal semplice fatto che non può esistere trasmissione della conoscenza e della memoria, che non implichi al contempo, variabili dosi di oblio: la memoria, anche quella documentaria, è per sua natura selettiva. Di questo son sempre stati ben consapevoli gli archivisti e lo sono soprattutto quelli contemporanei, dato che le dimensioni degli archivi con cui hanno a che fare hanno imposto forme di selezione talvolta molto drastiche (in molte realtà si distrugge attorno al 90% della documentazione prodotta, negli Stati Uniti si arriva fino al 97%). Il problema è ovviamente stabilire in base a quali criteri e con quali finalità la selezione viene operata. Problema che si pone anche nei confronti del Web, nel cui ambito attinge un livello di complessità probabilmente maggiore rispetto alla documentazione convenzionale. Infatti, seppure siano evidenti ripetizioni, futilità, ridondanze e banalità, c'è da dubitare che possano essere facil-

mente individuati, per i materiali che si accumulano sulla rete, criteri generali ed univoci, più o meno 'oggettivi', sulla cui base orientare scelte concrete di selezione.

2.2 *La conservazione del Web fra ambito nazionale e dimensione internazionale*

La conservazione del Web suscita in realtà interrogativi anche più profondi di quelli connessi alla quantità di documenti ed informazioni che vi circolano e alle tecnologie che possano garantire la sua 'archiviazione' e conservazione nel tempo. L'avvento di Internet e, più in generale, dei formati digitali, spinge piuttosto a domandarsi se e quanto i meccanismi istituzionali e le pratiche sociali, che hanno permesso fino adesso di tramandare alle generazioni future la memoria documentaria e il patrimonio di conoscenza e di cultura accumulati, siano ancora validi ed efficaci. Come è noto, la conservazione e la trasmissione della memoria documentaria è stata affidata, almeno nel corso degli ultimi due secoli, ad una rete di istituzioni (archivi nazionali o di Stato) la cui origine affonda nel processo di costruzione dello stato nazionale e nel ruolo che la storia ha esercitato all'interno di quel processo come fonte di legittimazione delle nuove realtà statali. Non dissimile sviluppo hanno avuto i meccanismi di conservazione e trasmissione della produzione editoriale, riservata ad istituzioni quali le biblioteche nazionali, che hanno acquisito molte delle loro caratteristiche a seguito dei processi di costruzione dello Stato nazionale, che le hanno plasmate da un lato come veri e propri 'archivi di libri', destinati a raccogliere il complesso della produzione intellettuale di una nazione (attraverso il deposito legale) e dall'altro come strumenti, che, attraverso la concentrazione di ciò che di più significativo si andava producendo a livello internazionale in ogni ambito dello scibile umano, dovevano porre la nazione nella condizione di tenere il passo e di competere con gli altri stati²².

È evidente che oggi questi meccanismi sociali ed istituzionali funzionano sempre meno per gli scopi per i quali si sono andati formando e consolidando nel corso del tempo. Le ragioni sono molteplici, ma certo l'avvento del digitale e soprattutto delle reti in questa crisi gioca un ruolo di primo piano. Internet, per i suoi caratteri di fenomeno globale che scavalca i confini territoriali, destabilizza e mette in primo

²² Sullo sviluppo e i caratteri delle biblioteche nazionali cfr. G. VITIELLO, *Alessandrie d'Europa. Storie e visioni di biblioteche nazionali*, Milano, Sylvestre Bonnard 2002.

luogo in crisi la dimensione nazionale sulla quale si è basata fino adesso la conservazione. Non solo. Le caratteristiche peculiari del Web – strumento di ‘pubblicazione editoriale’ più o meno tradizionale, ma anche di comunicazione fra le persone e fra istituzioni e cittadini – pongono anche altri interrogativi: ad esempio quali delle tradizionali istituzioni deputate alla conservazione e trasmissione del patrimonio culturale se ne deve occupare: e in che misura? le biblioteche, quelle nazionali in primo luogo? oppure anche gli archivi? oppure ancora – come sembra più logico – entrambe le istituzioni?

Ad interrogativi del genere non è facile dare risposte rapide ed efficaci, come sarebbe necessario. Lo dimostra bene, ad esempio, la difficoltà che hanno molte legislazioni nazionali sul deposito librario a prendere pienamente coscienza e ad adeguarsi alle novità introdotte dall’avvento della rete e dei supporti digitali e quando lo fanno non sempre le soluzioni adottate si dimostrano soddisfacenti²³. Ne è prova la legge approvata recentemente dal Parlamento italiano (la legge 106/2004)²⁴, che, se certo costituisce un passo avanti rispetto a quella precedentemente esistente, risulta fin dal titolo («Norme relative al deposito legale dei documenti di interesse culturale destinati all’uso pubblico») abbastanza vaga ed indeterminata nell’individuare i materiali che devono essere depositati, riferendosi ai «documenti destinati all’uso pubblico e fruibili mediante la lettura, l’ascolto e la visione, qualunque sia il loro processo tecnico di produzione, di edizione o di diffusione» ed affidandosi ad una elencazione che fa riferimento ora a tipologie di prodotti culturali (le carte geografiche, gli atlanti ecc.) ora – in particolare quando ci si riferisce al digitale – semplicemente ai supporti, prevedendo, genericamente, il deposito dei «documenti diffusi su supporto informatico» e dei «documenti diffusi tramite rete informatica» che non rientrano in nessuna delle categorie specificamente elencate. Categorie che, pensando a quanto circola sul Web, sembrano talmente indefinite che solo con l’approvazione del previsto regolamento applicativo sarà probabilmen-

²³ Per una analisi, in realtà non sempre aggiornata, sulla legislazione internazionale sul deposito legale cfr. la pagina ‘Legal Deposit’ del sito web della National Library of Australia *PADI – Preserving Access to Digital Information*, <<http://www.nla.gov.au/padi/topics/67.html>>.

²⁴ Legge 15 aprile 2004, n. 106: «Norme relative al deposito legale dei documenti di interesse culturale destinati all’uso pubblico», Gazzetta Ufficiale n. 98, 27 aprile 2004, <<http://www.parlamento.it/parlam/leggi/04106l.htm>>.

te possibile valutare l'effettiva portata della nuova normativa²⁵. Certo anche gli interrogativi suscitati da questa legge sembrano confermare quanto difficile sia tracciare, nel mondo digitale e nel Web in particolare, distinzioni nette fra le tipologie documentarie e quanto, allo stesso tempo, si vadano viepiù stemperando ed erodendo i confini fra compiti ed ambiti operativi (ad esempio quelli fra archivi e biblioteche) che nel 'mondo di carta' sembravano ovvi e scontati. Ma a dire il vero, tutto ciò è ampiamente dimostrato anche dai progetti, dalle esperienze e dalle realizzazioni pratiche che si vanno realizzando in questi anni proprio sul terreno della conservazione del Web.

3 Esperienze e progetti

3.1 L'Internet Archive

Gli antichi manoscritti della Biblioteca di Alessandria furono incendiati, gran parte dei primi prodotti della stampa non si salvarono e molti dei primi film furono riciclati per il loro contenuto d'argento. Mentre il World Wide Web di Internet non ha precedenti nel diffondere la voce popolare di milioni di persone che prima non sarebbe mai stata resa pubblica, nessuno provvedeva a conservare questi documenti e immagini (...) L'Internet Archive è un'organizzazione che sta raccogliendo i materiali ad accesso pubblico presenti su Internet per costruire una biblioteca digitale²⁶.

Con queste parole, uno dei pionieri di Internet, Brewster Kahle presentava nel fascicolo del marzo 1997 di «Scientific American» il progetto di 'archiviazione' dell'intero World Wide Web, o almeno della sua parte ad accesso libero e diretto, avviato l'anno prima dall'Internet Archive, organizzazione non-profit di San Francisco, e dava succintamente conto della 'filosofia', tutto sommato elementare e molto pragmatica, che aveva ispirato l'avvio dell'iniziativa e che da allora continua a guidarla,

²⁵ Sulla legge italiana sul deposito librario cfr. anche le valutazioni, in parte difforni da quelle formulate nel testo, del Gruppo di studio sulle biblioteche digitali dell'Associazione italiana biblioteche, *Nuova legge sul deposito legale e documenti digitali*, a cura di G. Bergamin, ultimo aggiornamento 24 giugno 2004, <<http://www.aib.it/aib/commiss/bdigit/depledig.htm>>.

²⁶ Cfr. B. KALE, *Preserving the Internet*, in «Scientific American», CCLXXVI, 3, March 1997, pp. 72-73.

nonostante la trasformazione di alcuni dei suoi caratteri iniziali, l'irrobustirsi del suo spessore problematico e la diversificazione dei suoi obiettivi. Assumendo come paradigma negativo la distruzione della biblioteca di Alessandria e la scomparsa di tanta parte della cultura classica e come ideologia positiva l'«universal access to human knowledge»²⁷ che sarebbe reso possibile dall'avvento di Internet, il progetto si propone di contrastare l'evanescenza del Web, con l'obiettivo di 'salvare' i siti che lo popolano e di conservare così testimonianza di Internet come fenomeno storico complessivo, a beneficio dei contemporanei e soprattutto delle generazioni future. L'intento dell'Internet Archive è di 'catturare' il maggior numero possibile di siti, senza alcun criterio selettivo predeterminato e prescindendo dal contenuto più o meno effimero che i siti stessi possono veicolare. Come ha dichiarato recentemente Kahle,

l'aspetto cruciale di collezioni bibliotecarie non selettive è che non puoi dire in anticipo ciò che risulterà importante. Il Web è il medium della gente, non è elitario. Ciascuno può pubblicarvi, così ci trovi il buono, il cattivo, lo sgradevole, il blasfemo. È come siamo noi, questa è la cosa avvincente [...]²⁸.

Questo mastodontico progetto di conservazione generalizzata del Web è perseguito grazie ai *crawler* di una compagnia privata specializzata nello sviluppo di sistemi di navigazione e di analisi dei siti Internet, la Alexa Internet, fondata dallo stesso Kahle e poi venduta ad Amazon. I robot di Alexa passano in rassegna i siti accessibili sull'intero Web e ne memorizzano il contenuto, producendo una sorta di istantanea (*snapshot*, come viene definita in linguaggio tecnico) di ciascun sito al momento della sua cattura. Dopo sei mesi, il materiale memorizzato è ceduto gratuitamente all'Internet Archive. Il ritmo con cui i siti web vengono catturati è variabile. Mentre nei primi tempi di esistenza dell'Internet Archive venivano effettuate, in media, una decina di istantanee ogni anno, oggi esse sono anche più di una alla settimana.

²⁷ È uno dei motti che campeggia nelle pagine del sito dell'Internet Archive, <<http://www.archive.org>>. La visione di fondo che ispira il progetto è stata esposta sinteticamente in *A Conversation with Brewster Kahle*, in «ACM Queue», II, 4 giugno 2004, <<http://www.acmqueue.com/modules.php?name=Content&pa=showpage&pid=163&page=1>>.

²⁸ Cfr *Way back when...*, intervista di P. Marks a Brewster Kahle, in «New Scientist», CLXXVI, 2370, 23 novembre 2003, pp. 46-49.

Dal 1996 al momento in cui scriviamo è stato assemblato un archivio di circa 55 miliardi di pagine per un totale approssimativo di 1 petabyte che cresce con la media di 20 terabyte al mese, che equivale all'intera Library of Congress²⁹. Del proprio modello ispiratore, la biblioteca pubblica di matrice anglosassone, l'Internet Archive si propone di replicare la libera e generalizzata accessibilità ai materiali conservati. «L'accesso libero e gratuito alla letteratura e agli altri scritti – ribadisce il sito del progetto – è stato da lungo tempo considerato essenziale all'educazione e alla conservazione di una società aperta», per questo l'Internet Archive si propone di «offrire accesso permanente» alle proprie collezioni in formato digitale a storici, a ricercatori e al pubblico in generale³⁰.

Sviluppando una prospettiva del genere, l'Internet Archive è andato negli ultimi anni estendendo la propria strategia di raccolta di materiali digitali anche oltre il Web, includendo nelle proprie collezioni audiovisivi, registrazioni sonore e testi digitali. Ha inoltre differenziato i propri servizi di archiviazione del Web, sviluppando una sorta di *Web archiving on demand*, che permette a singole istituzioni di portare avanti progetti di conservazione dei propri siti e/o di altri siti web centrati su specifici siti o su determinate tematiche o eventi³¹.

Nell'ottobre 2001 l'archivio è stato reso consultabile da remoto attraverso un apposito motore di ricerca, la Wayback Machine, che consente il recupero e la visualizzazione dei siti web archiviati sulla base della loro URL e della data di memorizzazione.

Il nome del motore di ricerca sembra molto azzeccatto. In effetti, più che uno strumento di ricerca di un archivio organico – o quanto meno organizzato secondo una logica riconoscibile – Wayback Machine richiama alla mente una scatola di vecchie fotografie disordinate, che danno la sensazione di viaggiare indietro nel tempo e riportano alla mente il ricordo sfuocato di un passato più o meno lontano. Essa fa scoprire l'indubbio fascino di visitare siti che non esistono più o che sono

²⁹ Per le cifre riportate nel testo cfr. 'Internet Archive Frequently Asked Questions', <<http://www.archive.org/about/faqs.php>>.

³⁰ 'About the Internet Archive', all'URL <<http://www.archive.org/about/about.php>>.

³¹ «Il nuovo servizio per sottoscrizione dell'Internet Archive, Archive-It permette a singole istituzioni di costruire, gestire e ricercare il proprio archivio di siti web, per mezzo di un applicativo Web user friendly, senza richiedere nessuna particolare competenza tecnica: 'Welcome to Archive-It', <<http://www.archive-it.org/index.php>>.

nel frattempo radicalmente mutati – ribaltando la logica effimera del Web. Ma come le vecchie fotografie disordinate, un po' sgualcite e con qualche strappo che ci priva di particolari importanti, anche le pagine web alle quali la Wayback Machine permette di accedere danno talvolta l'idea di relitti conservati un po' a caso e senza tutta la cura che sarebbe stata necessaria³².

Non è raro che i siti 'catturati', soprattutto in determinati periodi, siano privi di immagini. Contenuti multimediali e siti interattivi non vengono restituiti propriamente. Link ipertestuali fra le pagine di uno stesso sito, anche fra semplici pagine html, talvolta non funzionano correttamente, mentre in altri casi rinviano automaticamente, non alle pagine del sito presenti nell'archivio come dovrebbero, ma a quelle ancora *online*. Non sono comprese nell'archivio pagine che si formano dinamicamente *on the fly* o il cui recupero è possibile solo grazie alla compilazione di form (in sostanza quelle che compongono il *deep Web*), così come sono ovviamente esclusi siti o parti di siti il cui accesso è subordinato alla digitazione di *password* e ad altre restrizioni imposte dai gestori. Insomma il grado di completezza e d'integrità di molti siti è tutt'altro che ottimale, ma quasi mai l'utente è in grado di capire cosa manca e perché manca. Né esiste, oltre alla data di memorizzazione, alcun tentativo, anche solo a livello elementare, di contestualizzare i siti, in relazione agli enti o alle persone che ne sono responsabili. Inoltre, l'assenza di selettività nella cattura dei siti non implica certo che l'archivio, pur essendo ricco di ridondanze e di duplicazioni, che secondo calcoli attendibili ascendono a circa il 30%³³, possa considerarsi esaustivo e privo di lacune. Implica semmai una certa casualità nelle presenze e delle assenze. Ma questo va da sé, data l'estensione del Web, i meccanismi che presiedono alla cattura dei siti e i caratteri dell'iniziativa. Per l'utente resta tuttavia difficile valutare la reale rilevanza del materiale disponibile rispetto alla totalità di quello a suo tempo pubblicato sui siti. Talvolta assenze significative emergono facilmente. Capita così di trovare 'immagini' di un determinato sito, ma non per le date che ci aspetterebbe.

Infine, resta aperto l'interrogativo sulla reale possibilità di conservazione a lungo termine di un archivio digitale dalle dimensioni straordi-

³² Per un'analisi dei problemi che i materiali dell'Internet Archive presentano cfr. anche DAY, *Collecting and Preserving the World Wide Web* cit.

³³ ROBERTS, *Inside the Internet Archive* cit.

narie come questo. Appare indubbio che il gigantismo dell'impresa, che non influisce positivamente sulla qualità delle presentazioni dei singoli siti, nonché sulla loro completezza e contestualizzazione, è destinato anche a provocare, con il passar del tempo, problemi di uguali dimensioni proprio dal punto di vista della conservazione e accessibilità.

Pur con i limiti segnalati, l'Internet Archive costituisce a tutt'oggi una risorsa indubbiamente importante, probabilmente unica, non solo per tracciare la storia del Web nel suo complesso o di suoi singoli segmenti, ma anche per attingere a materiali capaci di illustrare gli infiniti aspetti della società e del mondo contemporanei che, nel passaggio dal secondo al terzo millennio, sono stati documentati da siti Internet attualmente non più accessibili o profondamente trasformati.

3.2 *Il September 11 Archive*

C'è soprattutto un evento che è doviziosamente documentato non tanto nell'Internet Archive come tale ma in un progetto parallelo condotto a suo tempo con una logica un po' diversa da quella dell'iniziativa di Bewster Kahle, ma che ad essa è stato poi associato, ed è l'attacco suicida alle Torri gemelle e gli altri avvenimenti dell'11 settembre 2001.

Il 'September 11 archive' è una raccolta di siti web ad un tempo tematica (materiali che hanno attinenza con l'attacco alle Twin Towers e con gli altri contestuali episodi terroristici nonché con le reazioni che questi episodi hanno determinato) e circoscritta cronologicamente. Essa è il risultato dell'iniziativa di Kirsten Foot, professore di scienza delle comunicazioni all'Università di Washington e di Steven Schneider, professore di scienze politiche presso l'Università dello Stato di New York che, coadiuvati da alcuni volontari e in collaborazione con l'Internet Archive e con la Library of Congress di Washington, hanno 'catturato' e memorizzato migliaia di siti web all'indomani dell'11 settembre³⁴.

³⁴ La Library of Congress, in collaborazione con l'Internet Archive e con l'organizzazione promossa da Foot e Schneider, WebArchivist.org, ha promosso altre raccolte di siti web tematiche oppure legate ad eventi specifici, come ad esempio le elezioni presidenziali del 2000 e quelle di mezzo termine del 2002; ad esse si può accedere dall'home page del progetto Minerva, promosso dalla Library of Congress <<http://www.loc.gov/minerva/index.html>>. Sui caratteri di questi progetti e i problemi posti dalla loro realizzazione cfr. S.M. SCHNEIDER et al., *Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and the Election 2002 Web Archivi*, in «3rd ECDL Workshops on Web Archiving», Trondheim (Norvegia), 21 agosto 2003, <

L'aggregazione di questo archivio è paradigmatica da più punti di vista. Innanzi tutto perché essa è il risultato di una consapevolezza, fino allora scarsamente diffusa, del ruolo cruciale giocato ormai dal Web nel documentare avvenimenti e fenomeni del mondo contemporaneo, grazie al quale vengono sedimentandosi fonti potenziali di estremo interesse per scienziati sociali e storici. Da questo punto di vista l'11 settembre ha certamente segnato una svolta, testimoniata anche dall'interesse che l'iniziativa ha suscitato anche presso gli organi di comunicazione di massa, come stampa e televisione³⁵.

Non c'è dubbio, poi, che i materiali raccolti abbiano uno speciale interesse in sé, proprio per il fatto che l'attentato alle Torri Gemelle è stato un evento di rilievo mondiale, che ha avuto in Internet ripercussioni così ampie e ramificate come mai prima di allora era probabilmente accaduto. L'archivio che ne raccoglie le testimonianze si presta perciò efficacemente ad una verifica di quali e quanti elementi di novità e di specificità le reti telematiche introducano, rispetto ad altri media di massa, come la radio e la televisione, nei processi di costruzione degli eventi e nelle forme della loro ricezione da parte di individui e comunità. Come hanno dimostrato i primi studi condotti sull'archivio³⁶, vivere un evento attraverso Internet significa non solo assistere al suo dispiegarsi – come avviene anche con le dirette televisive e radiofoniche – ma significa dare un contributo in prima persona a definirlo come tale, attraverso la propria diretta testimonianza, la manifestazione delle proprie reazioni, la discussione di quelle degli altri, la proposta di proprie interpretazioni

bnum.bnf.fr/ecdl/2003/proceedings.php?f=schneider>. L'archiviazione dei siti web su base tematica è stata condotta dall'Internet Archive anche per più recenti eventi, quali ad esempio l'uragano Katrina ed è fra le possibilità offerte con il servizio di *Web archiving on demand*, Archive-it. La raccolta su Katrina è accessibile all'URL <<http://Websearch.archive.org/katrina>>.

³⁵ Su Internet e gli avvenimenti dell'11 settembre 2001, cfr. l'ampia e documentata analisi di G. RONCAGLIA, *Internet e l'11 settembre: un caso di studio*, in M. CALVO et al., *Internet 2004* cit., anche online, <<http://www.merzWeb.com/testi/saggi/11settembre.htm>>.

³⁶ Per una prima analisi di come gli avvenimenti dell'11 settembre si sono rispecchiati in Internet cfr. i rapporti del PEW INTERNET & AMERICAN LIFE PROJECT, *The Commons of the Tragedy: How the Internet was used by millions after the terror attacks to grieve, console, share news, and debate the country's response*, 10 ottobre 2001, <<http://www.pewinternet.org/reports/toc.asp?Report=46>>; PEW INTERNET & AMERICAN LIFE PROJECT, *One year later: September 11 and the Internet*, 5 settembre, 2002, <<http://www.pewinternet.org/reports/toc.asp?Report=69>>.

e l'ascolto di quelle degli interlocutori. Internet si propone insomma come uno strumento attraverso il quale i singoli individui possono contribuire ad elaborare il senso e la portata collettiva degli avvenimenti. I materiali che questo processo sedimenta, possono perciò offrire un quadro – molto più sfaccettato e multiforme di quello restituito dai media tradizionali – delle risposte individuali e collettive, uno spaccato, a grana quasi molecolare, e allo stesso tempo di massa, delle reazioni di consistenti spaccati della società di fronte ad eventi significativi, di dimensioni locali o globali.

Come si è accennato precedentemente, le modalità di costituzione del 'September 11 Archive' hanno seguito percorsi diversi da quelli adottati in genere nell'Internet Archive. Come spiegano i responsabili dell'iniziativa nelle pagine di presentazione del sito web che dà accesso all'archivio:

i nostri sforzi iniziali identificarono circa 30.000 differenti siti. Ciascuno dei siti identificati fu quotidianamente archiviato fino al primo dicembre 2001. L'obiettivo dell'attività di archiviazione fu di conservare non solo i bit e il contenuto, ma anche la dimensione di esperienza diretta prodotta da questa Web sfera che emergeva rapidamente. Catturando pagine e siti con il loro contesto di link, gli strumenti di archiviazione hanno conservato non solo la collezione delle pagine web, ma anche la sfera Web connessa attraverso link – caratterizzata e tenuta insieme dall'orientamento verso un oggetto comune o da un medesimo punto di riferimento, in questo caso gli attacchi dell'11 settembre. L'archivio nel suo complesso consiste di 5 terabyte di materiali Web³⁷.

Su questi materiali è in corso un'opera di catalogazione che finora ha interessato circa il 10% della raccolta e che mira ad elaborare descrizioni per il pubblico dotate di una certa accuratezza ed a fornire informazioni 'di contesto' sui responsabili dei siti, che «dovrebbero aggiungere una valore [aggiunto] significativo per gli utenti dell'archivio». Si è, ad esempio, cercato di classificare i 'soggetti produttori' sulla base di una griglia tipologica predefinita, che può essere utilizzata anche in sede di consultazione dell'archivio. C'è in insomma il tentativo di fornire non soltanto dei materiali, ma un insieme di documentazione presentato in modo tale che appaia dotato di un certa organicità³⁸.

³⁷ 'Welcome to the September 11 Web Archive', all'URL <<http://september11.archive.org/welcome.html>>.

³⁸ Per i criteri di descrizione adottati che fanno riferimento alle pratiche di catalogazione bibliografica, ed in particolare a quelle della Library of Congress, cfr. SCHNEIDER *et*

3.3 Progetti di agenzie nazionali

L'alternativa fra 'cattura' indifferenziata e selezione si ripropone anche per progetti di 'archiviazione' del Web che si muovono in ambiti più ridotti di quelli dell'Internet Archive e che sono portati avanti soprattutto dalle biblioteche nazionali di vari paesi, ad integrazione di quei compiti di raccolta e conservazione del patrimonio bibliografico della nazione ad esse affidato.

Si sono orientate verso la 'cattura' generalizzata del proprio spazio Web nazionale le biblioteche nazionali dei paesi scandinavi, come quella finlandese e quella svedese. Quest'ultima, in particolare, ha avviato fin dal 1997 un progetto di archiviazione dei siti web, riferibili alla Svezia. L'obiettivo già di per sé pone qualche problema, poiché se è abbastanza semplice individuare i siti registrati sotto il dominio di primo livello '.se', un po' più complicato è scovare quelli che possono considerarsi svedesi, ma che sono registrati sotto altri domini nazionali di primo livello, oppure sotto domini internazionali, quali '.com', '.org', o '.net'. La raccolta, condotta per mezzo di *crawler* che percorrono lo spazio Web svedese due o tre volte l'anno, è cresciuta molto rapidamente. Nel 1997 furono memorizzate 6 milioni e ottocentomila pagine provenienti da 15.700 siti; nel 2001 i siti erano saliti a 126.000, per un totale di circa 30 milioni di file. Utilizzando tecnologie di *harvesting* non dissimili da quelle dell'Internet Archive, il progetto svedese incontra difficoltà analoghe dal punto di vista della completezza e dell'integrità dei siti catturati³⁹.

Adotta invece strategie severamente selettive il progetto PANDORA (*Preserving and Accessing Networked Documentary Resources of Australia*) messo a punto, a partire dal 1996-1997, della National Library of Australia⁴⁰. Le ragioni di un tale approccio sono state chiaramente indi-

al., *Building Thematic Web Collections* cit.

³⁹ Sui progetti di conservazione dello spazio Web svedese cfr. A. ARVIDSON-K. PERSON-J. MANNERHEIM, *The Kulturarw3 Project – The Royal Swedish Web Archive: An example of 'Complete' Collection of Web Pages*, intervento alla LXVI Conferenza Generale dell'IFLA, Gerusalemme, 13-18 agosto 2000, <<http://www.ifla.org/IV/ifla66/papers/154-157e.htm>>; A. ARVIDSON, *The Collection of Swedish Web Pages at the Royal Library – The Web Heritage of Sweden*, intervento alla LXVIII Conferenza Generale dell'IFLA, Glasgow, 18-24 agosto 2002, <<http://www.ifla.org/IV/ifla68/papers/111-163e.pdf>>. Ai sensi della legge svedese l'accesso all'archivio è limitato unicamente alla sede della Biblioteca Nazionale.

⁴⁰ Una prima buona informazione sul progetto PANDORA si può leggere in E. PHIL-

cate da uno dei suoi responsabili:

primo, trattare le pubblicazioni *online* richiede molto lavoro e a maggior ragione è molto costoso. Secondo, per quanto suggerisce la nostra valutazione professionale, molte pubblicazioni *online* hanno uno scarso valore per la ricerca corrente e futura. Noi riteniamo che sia meglio identificare quei titoli che con maggiore probabilità siano destinati ad avere un interesse futuro per la ricerca e applicare le nostre limitate risorse a trattare questi ultimi, secondo *standard* elevati.

I siti da catturare sono scelti esplicitamente sulla base di alcuni criteri generali predefiniti, privilegiando quelli che «riguardano l'Australia; oppure sono scritti da un autore australiano e riguardano un argomento di significato e rilevanza sociale, politica, culturale, religiosa, scientifica o economica; oppure sono scritti da un australiano di riconosciuta autorità e costituiscono un contributo rilevante alla conoscenza a livello internazionale»⁴¹. Ciò non toglie che si cerchi di documentare anche alcuni aspetti dell'uso 'sociale' del Web in ambito australiano, catturando siti, che più che per il loro specifico contenuto informativo, sono significativi per la popolarità o l'interesse che hanno suscitato. Concentrandosi su un numero tutto sommato ristretto di siti (nel gennaio 2006 il catalogo per l'accesso a PANDORA comprendeva circa 11000 'titoli' per un totale di 21500 snapshot dei siti e di circa 1100 gigabyte di memoria)⁴², il livello di qualità raggiunto nel trattamento delle loro 'immagini' memorizzate e nella loro catalogazione è indubbiamente elevato. Le procedure di cattura ed archiviazione prevedono, fra l'altro, la memorizzazione di una copia

LIPS, *Ensuring Long-Term Access to Online Publications*, in «The Journal of Electronic Publishing», IV, 4 giugno 1999, <<http://www.press.umich.edu/jep/04-04/phillips.html>>, dal quale sono tratte le citazioni seguenti, se non diversamente indicato. L'home page del sito web del progetto, dalla quale si può accedere alla consultazione dell'archivio e a molti materiali informativi, è <<http://pandora.nla.gov.au/index.html>>; cfr. anche P. KOERBIN, *The Pandora Digital Archiving System (Pandas) and Managing Web Archiving in Australia: A Case Study*, in «4th International Web Archiving Workshop», Bath (UK), 16 settembre 2004, <<http://www.iwaw.net/04/proceedings.php?f=Koerbin>>.

⁴¹ Cfr. le «Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia», <<http://pandora.nla.gov.au/selectionguidelines.html>>.

⁴² Cfr. 'PANDORA archive size and monthly growth. Statistics as at 15 January 2006', <<http://pandora.nla.gov.au/statistics.html>>.

master che sia identica in ogni suo aspetto a quella sul sito originario, sia in termini di contenuto che di apparenza esteriore e di sensazione che produce. Quando per ragioni tecniche alcune funzionalità del sito originario non sono riproducibili (ad esempio motori o altri strumenti di ricerca), si cerca di preservare almeno il contenuto intellettuale nella sua interezza. Quando la versione da porre in consultazione è stata confezionata il responsabile del sito originario è invitato a visionarla e ad esprimere eventuali commenti. Uno dei punti di forza del progetto è infatti lo sforzo di instaurare rapporti di collaborazione con i soggetti responsabili dei siti, ai quali, in mancanza di normative definite sul deposito legale delle pubblicazioni *online*, viene richiesto il consenso all'acquisizione del sito da parte della biblioteca e, talvolta, l'invio dei file da conservare nell'archivio.

I siti 'archiviati' sono tutti catalogati secondo i tradizionali standard bibliografici e inclusi sia nella bibliografia nazionale australiana che nell'OPAC della biblioteca. Oltre che attraverso questi strumenti, ad essi è possibile accedere anche grazie a due liste, l'una alfabetica, l'altra articolata per grandi aree tematiche. Selezionando una voce delle liste non si accede direttamente al sito, ma ad una scheda che ne fornisce alcuni elementi di descrizione e di contestualizzazione, che, nell'intenzione dei responsabili, in futuro dovrebbero essere ulteriormente arricchiti, soprattutto grazie alla predisposizione di un sistema di metadati che dovrebbe consentire efficaci ricerche all'interno dei materiali che compongono ciascuno dei siti conservati.

Lo stesso approccio alla selezione dei siti da archiviare, la stessa tecnologia, cioè il sistema PANDORA, ed i medesimi criteri di descrizione e di presentazione dei siti messi a punto dalla National Library of Australia sono stati adottati anche da un consorzio di istituzioni culturali britanniche, fra le quali la British Library, la National Library of Wales, la National Library of Scotland e The National Archives. I siti web da archiviare, segnalati dalla diverse istituzioni che fanno parte del consorzio, sono 'memorizzati' e conservati da un unico centro. Essi sono accessibili nel loro insieme a partire da un unico sito, all'interno del quale la ricerca può essere condotta scorrendo una lista alfabetica delle denominazioni dei siti, selezionando una voce in uno schema di categorie e subcategorie tematiche oppure interrogando un motore di ricerca⁴³. Inoltre, ciascuna delle istituzioni partner del progetto sta prov-

⁴³ Cfr. <<http://www.Webarchive.org.uk>>. L'archivio è in linea dal maggio 2005.

vedendo ad inserire nei propri cataloghi la descrizione dei siti di cui è responsabile. Una tale metodologia di lavoro ha permesso di far confluire all'interno di un unico progetto interessi e strategie conservative delle istituzioni partecipanti, salvaguardando al contempo le specificità e le sensibilità di ciascuna di esse⁴⁴.

The National Archives of United Kingdom (nel quale sono confluiti recentemente alcune istituzioni archivistiche britanniche fra le quali il Public Record Office di Londra) ha anche avviato, a partire dall'agosto 2003, un proprio progetto di cattura e archiviazione di alcuni siti del governo inglese, della conservazione dei cui archivi è responsabile per mandato istituzionale. Anche in questo caso il criterio è fortemente selettivo, poiché come ha spiegato uno dei suoi responsabili «non sarebbe utile per gli utenti o effettivamente realizzabile l'acquisizione di tutti i siti web governativi». Sono stati così identificati una

cinquantina di siti principali che coprono tutte le attività più importanti del governo e le responsabilità dei Ministri che fanno parte del Gabinetto. I temi identificati riguardano: dipartimenti centrali e amministrazioni delegate, difesa e politica estera, amministrazione della giustizia e sicurezza interna, gestione delle risorse, salute, educazione e cultura e, infine, regolamentazione e coordinamento degli affari [...] La maggior parte [dei siti] sarà catturata ogni sei mesi, ma alcuni di essi (Foreign Office, Defence and International Development) lo saranno settimanalmente. La nostra intenzione è di essere flessibili: se una tema di una certa importanza emerge, allora risponderemo acquisendo il sito governativo pertinente⁴⁵.

In una prima fase è stata sperimentata l'acquisizione dei *backup* dei siti direttamente dai loro responsabili. Successivamente, a causa delle difficoltà pratiche incontrate con i *backup* pervenuti, è stato deciso di affidare la realizzazione pratica del progetto alle tecnologie e alle metodologie adottate dall'Internet Archive di San Francisco. Anche l'accesso alle diverse versioni dei siti catturati avviene attraverso la Wayback machine. Tramite il sito dei National Archives britannici è così oggi possi-

⁴⁴ S. BAILEY-D. THOMPSON, UKWAC: *Building the UK's First Public Web Archive*, in «D-Lib Magazine, Digital Libraries, Digital Library Research», XII, 1 gennaio 2006, <<http://www.dlib.org/dlib/january06/thompson/01thompson.html>>.

⁴⁵ D. THOMAS, *Digital Preservation at The National Archives*, <<http://www.nationalarchives.gov.uk/preservation/digitalarchive/pdf/dpattna.pdf>>.

bile avere una visione diacronica dell'evoluzione nel corso degli ultimi due anni dei siti web del governo britannico, alcuni dei quali rivestono un'importanza cruciale poiché sono stati il mezzo di comunicazione politica quasi esclusivo utilizzato da alcune di quelle istituzioni: basti pensare al ruolo che ha avuto il sito web nel pubblicizzare gli sviluppi e i risultati del lavoro dell'«Hutton Inquiry», cioè della Commissione presieduta da Lord Hutton, incaricata di investigare sulle circostanze della morte di David Kelly⁴⁶.

Altre iniziative intraprese o progettate da alcune biblioteche nazionali si vanno orientando verso una soluzione intermedia fra la cattura indifferenziata dei siti, attraverso il loro periodico *download* per mezzo di *crawler*, ed una selezione più mirata, che, per essere più efficace, ricorre anche alla collaborazione dei responsabili dei siti stessi. Tali iniziative si propongono di mettere a frutto gli esperimenti condotti dalla Bibliothèque Nationale de France⁴⁷ e dal progetto NEDLIB (Networked European Deposit Library) finanziato dalla Comunità Europea qualche anno fa⁴⁸. Tali iniziative hanno recentemente condotto alla costituzione di un consorzio internazionale per l'archiviazione del Web al quale hanno aderito, oltre all'Internet Archive, le biblioteche nazionali di alcuni paesi europei ed extraeuropei⁴⁹. Il Consorzio si propone di

⁴⁶ I siti web del governo britannico possono essere consultati, insieme ad informazioni sulla metodologia della loro raccolta, a partire dall'indirizzo <<http://www.nationalarchives.gov.uk/preservation/Webarchive/>>.

⁴⁷ Per i progetti e le iniziative della Bibliothèque Nationale de France cfr. S. ABITEBOUL *et al.*, *A First Experience in Archiving the French Web*, in «SpringerLink: Lecture Notes in Computer Science 2458», 13 settembre 2002, <<http://link.springer.de/link/service/series/0558/papers/2458/24580001.pdf>>; J. MASANÈS, *Towards Continuous Web Archiving: First Results and an Agenda for the Future*, in «D-Lib Magazine», VIII, 12, December 2002, <<http://www.dlib.org/dlib/december02/masanès/12masanès.html>>; ID., *Archiving the Deep Web*, presentazione power point, 2nd European Conference on Digital Libraries Workshop on Web Archiving, Roma, 19 settembre 2002, <<http://bibnum.bnf.fr/ecdl/2002/BnF/BnF.html>>.

⁴⁸ Cfr. J. HAKALA, *Collecting and Preserving the Web: Developing and Testing the NEDLIB Harvester*, in «RLG DigiNews» V, 2, <<http://www.rlg.org/legacy/preserv/diginews/diginews5-2.html>>.

⁴⁹ I membri dell'International Internet Preservation Consortium, oltre alla Bibliothèque Nationale de France e all'Internet Archive, sono la Biblioteca Nazionale Centrale di Firenze, Det Kongelige Bibliotek danese, Helsingin yliopiston kirjasto - Suomen

sviluppare la collaborazione internazionale nell'ambito delle politiche di conservazione a lungo termine del Web e nella messa a punto degli strumenti tecnologici, per «rendere possibile la raccolta di una significativa quantità dei materiali che circolano in Internet [...] in modo tale che possano essere conservati, salvaguardati e resi accessibili nel lungo periodo»⁵⁰.

Secondo i loro promotori per avere un archivio Web di buona qualità e rappresentatività dovrebbero essere introdotti alcuni meccanismi di correzione degli automatismi dei *crawler*, che siano in grado di tener conto in maniera efficace delle specifiche dinamiche e dei ritmi di aggiornamento dei siti interessati.

La frequenza di 'cattura' dovrebbe essere quindi commisurata alle caratteristiche dei singoli siti e alla loro importanza relativa.

La dimensione internazionale di questa iniziativa mette bene in luce la crescente destabilizzazione dei tradizionali meccanismi di conservazione del patrimonio culturale fondati su basi nazionali, che l'avvento di Internet ha provocato. D'altronde quanto del Web degli anni appena trascorsi si è salvato, – quel poco o quel molto è ancora difficile dirlo – è in buona parte il risultato dell'iniziativa di un soggetto del tutto atipico come l'Internet Archive che, come si è visto, opera su un orizzonte decisamente 'globale'.

4 *Per un modello alternativo di archiviazione del Web*

4.1 *I caratteri (e i limiti) dei progetti in corso*

Come si è cercato di mostrare, intorno al problema della conservazione del Web si va sviluppando un certo numero di progetti di rilievo nazionale od internazionale, basati su modelli concettuali e su scelte tecniche spesso alquanto difforni. L'assenza di univocità nelle strategie conservative non va considerata soltanto come un indizio delle difficoltà che la soluzione del problema comporta né delle attuali insufficienze teoriche o della persistente immaturità delle tecnologie – elementi certo in parte presenti. Né è necessariamente vero che, in prospettiva, debba

Kansalliskirjasto, Kungliga biblioteket Sveriges nationalbibliotek, Landsbokasafn Islands – Haskolabokasafn, Library and Archives Canada, Nasjonalbiblioteket norvegese, National Library of Australia, The British Library, The Library of Congress.

⁵⁰ Ulteriori informazioni sul sito web del progetto, <<http://www.netpreserve.org/>>.

afferinarsi un unico modello conservativo. La complessità del Web e la varietà di motivazioni che possono ispirarne l'archiviazione sono ragioni più che sufficienti a far sì che anche i progetti che verranno messi in campo in un futuro più o meno prossimo si possano distribuire all'interno di uno spettro più o meno ampio di possibilità. Né ciò deve essere considerato come un fatto necessariamente negativo. Al contrario, l'esistenza di una pluralità di progetti, in parte alternativi, in parte coincidenti, può dare alla fragile memoria del Web qualche *chance* in più di sopravvivere.

Ma nonostante talune diversità, i progetti esaminati hanno in comune molti caratteri. Due sono quelli che sembrano soprattutto qualificanti. Da un lato, l'affidarsi largamente a modelli concettuali e a strumenti tecnologici che sono stati originariamente messi a punto per esplorare e 'catalogare' le risorse *online*, e non certo per promuovere una sedimentazione organica della memoria di Internet, quali sono appunto le tecnologie dei *crawler* sviluppate per la ricerca e l'indicizzazione delle risorse Web; dall'altro il considerare i siti e il Web nel suo complesso come nuove tipologie di 'pubblicazioni', da trattare secondo i parametri, pur aggiornati, delle discipline bibliografiche tradizionali – approccio che è soprattutto tipico dei progetti elaborati e condotti dalle biblioteche nazionali, ma che è presente, seppure in forme più confuse, anche nell'Internet Archive.

Che sul Web esistano proiezioni dei prodotti editoriali esistenti, per così dire, nel mondo di carta i quali, dei propri modelli originari, conservano molte caratteristiche 'genetiche', è un fatto difficilmente contestabile, basti pensare ai periodici elettronici, alle raccolte di testi delle biblioteche digitali ecc. Se tutto ciò è certamente vero, è tuttavia altrettanto vero che il Web è un fenomeno molto più complesso, difficilmente riducibile, nel suo insieme e in molte delle sue singole componenti, ad un prodotto o ad un insieme di prodotti tradizionalmente bibliografici, facilmente classificabili con i tradizionali strumenti offerti dalle discipline biblioteconomiche. Ci sono nel Web anche dimensioni di altro genere. C'è ad esempio una dimensione tipicamente archivistica, che si presenta sotto forme diverse, alcune delle quali tipiche di alcune risorse o di alcune tipologie di siti, altre, invece, più generalmente rintracciabili nel Web nel suo complesso.

4.2 Uno sguardo 'archivistico' sul Web

Innanzitutto, sul Web si accumulano archivi o parti di archivi, intesi nel senso più tradizionale del termine, di sedimenti documentari dell'attività pratica di un determinato soggetto produttore. Abbiamo ricordato

le esperienze di *e-government* e quelle delle reti civiche, all'interno delle quali i materiali che si aggregano sul Web, hanno strette relazioni con quelli che confluiscono all'interno degli archivi (cartacei od elettronici che siano) delle relative istituzioni promotrici, fino a configurarsi come una componente essenziale di questi. Ma una tendenziale identificazione fra archivio e spazio Web è rilevabile anche in altri contesti, ad esempio in quello dei movimenti che utilizzano il Web ed Internet nel suo complesso come strumento ad un tempo di organizzazione e di comunicazione. In questi casi, anzi, i materiali prodotti per e diffusi attraverso il Web costituiscono spesso le uniche tracce documentarie che movimenti e organizzazioni lasciano. Non è certo un caso che, soprattutto nei confronti del primo tipo di materiali – quelli cioè prodotti da istituzioni statali o comunque pubbliche – taluni archivi nazionali comincino a mostrare attenzione e a formulare prime ipotesi di strategie conservative. Ad esempio i National Archives di Washington hanno realizzato un intelligente recupero delle diverse versioni del sito web della Casa Bianca durante la presidenza Clinton⁵¹. Altre istituzioni archivistiche, quali ad esempio i National Archives of Australia, hanno emanato istruzioni e linee guida sulla gestione e conservazione nel tempo dei siti web delle istituzioni governative che ad essi fanno riferimento⁵².

L'approccio all'archiviazione dei siti web promosso e incoraggiato dalle istituzioni archivistiche è sensibilmente diverso rispetto a quello che abbiamo visto prevalere nella gran parte delle esperienze e dei progetti in corso. Mentre questi ultimi si basano largamente su meccanismi automatici di cattura periodica dei siti web da parte di istituzioni terze, le istruzioni e le linee guida elaborate in ambito archivistico, molto sensi-

⁵¹ Cfr. la sezione 'White House Web Sites during the Clinton Administration' (<<http://www.clintonlibrary.gov/archivesearch.html>>) del sito web: 'William Clinton Presidential Library' (<<http://www.clintonlibrary.gov/>>), nella cui presentazione si precisa: «La Casa Bianca ha conservato le 'immagini' di ciascuna versione [del proprio sito web]. Queste versioni sono materiali storici 'congelati nel tempo', il che significa che essi non sono più aggiornati e i link agli altri siti non funzionano».

⁵² Cfr. NATIONAL ARCHIVES OF AUSTRALIA, *Archiving Web Resources: a Policy for Keeping Records of Web-based Activity in the Commonwealth Government*, gennaio 2001 <http://www.naa.gov.au/recordkeeping/er/Web_records/archWeb_policy.pdf>; NATIONAL ARCHIVES OF AUSTRALIA, *Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government*, marzo 2001, <http://www.naa.gov.au/recordkeeping/er/Web_records/archWeb_guide.pdf>.

bili ai problemi della salvaguardia nel corso del tempo dell'affidabilità, dell'accuratezza e della autenticità dei materiali digitali, si propongono di rendere gli stessi soggetti responsabili dei siti protagonisti in prima persona di organiche strategie per la loro conservazione nel tempo. Essi si fanno cioè portatori di un modello che è, nella sostanza, quello di estendere in modo consapevole anche ai siti web le usuali politiche di gestione degli archivi; politiche che vedono, di solito, i loro soggetti produttori farsi carico della loro ordinata sedimentazione, selezione e conservazione fino al momento del loro versamento nelle istituzioni deputate a conservarli a beneficio delle future generazioni, cioè negli archivi storici. Solo così è possibile anche conservare in maniera adeguata le connessioni dei siti web con i sistemi documentari dei loro soggetti responsabili⁵³.

In realtà, a ben guardare, se alcuni siti web possono costituire una vera e propria componente di questi sistemi, data la funzione cruciale che svolgono nell'organizzare i flussi di lavoro e di comunicazione, una dimensione genericamente archivistica la possiedono tutti i siti web; o meglio tutti i siti web possono prestarsi ad essere proficuamente interpretati e gestiti secondo categorie che sono tipicamente archivistiche, proprio perché essi si sviluppano secondo dinamiche di accumulazione/sostituzione/aggiornamento dei materiali che si realizzano secondo un flusso temporale discreto, ma teoricamente ininterrotto, così come gli archivi si sedimentano nel corso del tempo senza soluzione di continuità⁵⁴. Perciò, come in una serie archivistica i singoli materiali acquistano significato in relazione a ciò che li precede e a ciò che li segue, così anche i materiali dei siti web, in una prospettiva storica, possono diventare fonti tanto più affidabili e significative, quanto più possono

⁵³ Per un approccio archivistico conservazione dei siti web, cfr. le considerazioni, scaturite nell'ambito del progetto internazionale InterPARES relativo alla conservazione a lungo termine degli archivi digitali, di J. SUNDERMAN, *Committing the Web to Memory: Transmitting Web-based Records over Time*, in *Il futuro della memoria: la trasmissione del patrimonio culturale nell'era digitale*, a cura di A. Spaziant, Torino, CSI-Piemonte 2005, pp. 217-224.

⁵⁴ Ha insistito sulla comparabilità fra il ciclo di vita dei siti web e quello degli archivi, scandito dalle fasi dell'archivio corrente, di deposito e storico, P. FELICIATI, *Il Web sugli archivi, gli archivi del Web*, intervento al convegno «Internet per il restauro, il restauro per Internet. Il Web per la conservazione, la conservazione del Web» (Ferrara, Salone del restauro, 3 aprile 2003), in «Kermes. La rivista del restauro», L, aprile-giugno 2003, pp. 61-63 (<http://www.minervaeurope.org/events/documents/Ferrara2003_Finale.pdf>).

essere criticamente collocati all'interno del flusso ininterrotto di evoluzione del sito. Prospettiva che, teoricamente, può essere pienamente conseguita solo ricorrendo alla conservazione dell'insieme delle condizioni che rendono possibile la completa ricostruzione dei siti «at any point of time» – come recita uno dei documenti dei National Archives of Australia, citato precedentemente⁵⁵.

Una strategia del genere – che non implica ovviamente la conservazione integrale dei materiali aggregatisi nei siti web, essendo una selezione razionale e lo scarto componente fondamentale di ogni politica conservativa⁵⁶ – non può essere in realtà realizzata senza il determinante concorso dei soggetti responsabili dei siti web, che soli sono in grado, nella predisposizione e nel mantenimento di essi, di adottare le soluzioni tecnologiche ed organizzative necessarie. D'altronde, ci sono componenti ed aspetti funzionali dei siti web, che non possono essere conservati altrimenti che dai loro responsabili. Già si è detto in precedenza di come il cosiddetto 'Web profondo', con le banche dati e i sistemi documentari che lo compongono, sfugga ai processi di cattura realizzati attraverso *crawler*. Anche la documentazione relativa alle transazioni fra gli utenti e il sito web, che riveste una particolare importanza quando questo è fortemente 'interattivo' o fornisce determinati servizi agli utenti, può essere acquisita e conservata solo per iniziativa dei soggetti responsabili dei siti, che soli possono predisporre una sua organica raccolta ed archiviazione. Si tratta di una documentazione che può in certi casi rivestire un'importanza cruciale, in prospettiva storica, perché può costituire testimonianza significativa per valutare l'efficacia di determinate politiche, del ruolo svolto in esse dai siti web e della loro capacità di configurarsi come strumento per instaurare rapporti fra utenti e soggetti responsabili dei siti, ad esempio fra cittadini e amministrazioni pubbliche. Proprio per questo – pur nel rispetto della privacy che richiede in casi come questi una cura particolare – ad essa dovrebbe essere dedicata una speciale attenzione dai responsabili dei siti, che oggi si limitano in genere a gestire semplicemente i file di log dei *server* Internet o a predisporre profili degli utenti da utilizzare a fini soltanto commerciali.

⁵⁵ Cfr. NATIONAL ARCHIVES OF AUSTRALIA, *Archiving Web Resources: a Policy for Keeping Records of Web-based Activity in the Commonwealth Government* cit.

⁵⁶ Sui problemi e le possibili strategie di selezione dei materiali presenti sui siti web cfr. J. LYLE, *Sampling the Umich.edu Domain*, in «4th International Web Archiving Workshop», Bath (UK), 16 settembre 2004, <<http://www.iwaw.net/04/proceedings.php?f=Lyle>>.

Infine solo una ‘archiviazione’ dei siti predisposta in modo razionale dai loro soggetti responsabili è in grado di documentare efficacemente il contesto giuridico, amministrativo, organizzativo e tecnico da cui i siti stessi hanno avuto origine e all’interno dei quali hanno operato. La documentazione del contesto alla lunga può dimostrarsi decisiva per la stessa comprensione dei materiali in essi aggregatisi nonché per una loro valutazione critica, insomma perché se ne possa fare un utilizzo proficuo come fonti storiche.

Insomma, che la fragile memoria del Web sopravviva in futuro in quantità rilevanti e in forme significative può dipendere non poco dalla volontà di tramandarla da parte degli enti, delle organizzazioni, delle persone che hanno contribuito a popolare il cyberspazio con i propri siti e dai comportamenti che costoro assumeranno nei suoi confronti⁵⁷. Tuttavia l’auspicabile diffondersi di una consapevolezza del genere, nonché delle strategie organizzative e delle soluzioni tecnologiche necessarie a metterla in pratica, non devono essere affatto considerate alternative rispetto ad iniziative di ‘archiviazione’ del Web promosse, sulla base di approcci metodologici diversi, da soggetti terzi, come l’Internet Archive, le biblioteche nazionali o anche, come è auspicabile, da istituzioni di dimensioni e raggio di azione più limitato, interessate alla conservazione di specifici spaccati, locali o tematici, del Web. Di fronte alla problematicità della conservazione a lungo termine, non solo del Web, ma dei materiali digitali in generale, la presenza di una molteplicità di progetti e di strategie non può non essere considerata – lo abbiamo già notato – che un fattore positivo, una possibilità in più che si sedimenti e venga trasmessa alle generazioni future una memoria documentaria digitale articolata e plurale, come è quella che si riflette negli archivi ‘di carta’ che il Novecento ci ha lasciato in eredità.

Un obiettivo che, in realtà, è ancora molto lontano da essere raggiunto.

STEFANO VITALI

⁵⁷ Sulla necessità che sia sviluppata una adeguata politica di conservazione dei siti web da parte dei loro responsabili insiste il *Manuale per la qualità dei siti web pubblici culturali, predisposto nell’ambito del progetto europeo Minerva dal gruppo di lavoro italiano* ‘Identificazione dei bisogni degli utenti e dei criteri di qualità per un accesso comune’. La «conservazione dei contenuti Web» costituisce una delle raccomandazioni formulate dal gruppo di lavoro ai fini della realizzazione di siti web di qualità: cfr. il capitolo 1.3 *Raccomandazioni generali: politiche e strategie* del *Manuale*, <<http://www.minervaeurope.org/publications/qualitycriteria-i/indice0402/capitoloprimo0402.htm#raccomandazioni>>.

Paesaggi di Informazione

1 Introduzione

In generale la composizione e la visualizzazione di informazioni testuali al calcolatore non sfrutta in modo completo tutte le *performance* grafiche disponibili nel sistema informatico.

Solitamente il testo scritto viene visualizzato nello schermo del calcolatore su uno sfondo bianco che simula la presenza di un foglio di carta da lettera delle dimensioni usuali. Il foglio digitale nel quale viene scritto il documento, presentato sul piano verticale della scrivania digitale, replica quindi la realistica disponibilità sullo schermo della pagina del documento che verrà stampata a compilazione ultimata.

La scrittura di un documento al calcolatore ricalca la stessa procedura utilizzata con le classiche macchine da scrivere dove il foglio di carta reale inserito tra i rulli a scorrimento meccanico è ora sostituito da uno spazio tridimensionale bianco posizionato sulla superficie dello schermo.

L'introduzione di sistemi operativi ad interfaccia utente ha permesso la disposizione di un elevato numero di 'finestre' o pagine bianche digitali sulla superficie dello schermo; solo una, però, delle tante pagine che l'utente può voler consultare contemporaneamente può essere di solito visualizzata completamente, mentre le altre risultano seminasconde o completamente occluse alla vista.

L'utente si trova quindi nella condizione di 'avvertire' la presenza di più pagine nascoste ma non è in grado di determinarne e verificarne il contenuto senza eseguire con il mouse operazioni di chiusura o spostamento delle finestre (o pagine) in primo piano (CHAPUIS 2007).

Per questo motivo la consultazione di più documenti digitali aperti contemporaneamente richiede operazioni della mano che controlla i movimenti del mouse paragonabili, e non più complesse, di quelle necessarie a ritrovare e riportare al centro del campo visivo e di lavoro documenti cartacei disposti in vario modo su una scrivania reale.

La stessa considerazione vale nel caso si desideri consultare un unico documento composto da un numero di pagine molto elevato: in questo

caso l'utente deve allenarsi nel fare eseguire alle applicazioni, sempre mediante il mouse, un movimento verticale continuo secondo la direzione di consultazione oppure, più semplicemente, azionare i comandi da tastiera relativi alla visualizzazione di una pagina dopo l'altra. Una tale procedura risulta semplice ma richiede comunque una azione continua da parte della mano dell'utente.

In questo modo quest'ultimo non ha una immediata percezione visiva del punto in cui si è posizionato rispetto alla totale estensione del documento se non andando a leggere il numero delle pagine.

A differenza di un documento o di un libro reale in cui le informazioni aptiche e visive riguardanti il peso, il volume o le dimensioni dello stesso, una volta che lo si è afferrato per la lettura, consentono di avere indicazioni dirette sul posizionamento relativo della pagina cercata rispetto all'inizio, in un sistema di trattamento di testi digitale ciò risulta meno diretto e richiede l'esecuzione di operazioni aggiuntive.

Naturalmente questa condizione risulta ancora aggravata quando il numero delle pagine del testo da consultare raggiunge valori molto elevati come avviene nel caso di opere quali, ad esempio, l'intera *Divina Commedia* o l'insieme delle opere di Shakespeare.

La gestione e visualizzazione al calcolatore di documenti testuali digitali viene quindi eseguita andando a sfruttare i comuni paradigmi di lettura utilizzati con il media convenzionale, cioè pagine di testo presentate verticalmente sullo schermo una dopo l'altra in un ambiente puramente bidimensionale.

L'utilizzo di sistemi basati su ipertesto consente di poter richiamare con facilità l'informazione memorizzata in altre aree, aumentando così l'efficienza nella disponibilità dell'informazione, ma continua ad utilizzare il paradigma di interazione della carta stampata. Ne risulta che le potenzialità di calcolo e di rappresentazioni dell'informazione propria dei calcolatori dell'ultima generazione non venga pienamente sfruttata.

Le presentazioni delle attuali schede grafiche consentirebbero invece di poter generare e gestire in tempo reale la visualizzazione di scenari tridimensionali anche complessi in cui, ad esempio, il numero di entità rappresentate e la loro intrinseca complessità geometrica risultino estremamente elevate.

Lo sviluppo, avvenuto nel corso dell'ultimo decennio del secolo scorso, di Ambienti Virtuali complessi che integrino elevate capacità di interazione realistica da parte dell'utente, ha consentito di identificare delle soluzioni innovative anche nel campo della visualizzazione di grandi moli di testo partendo alla creazione di particolari sistemi informatici denominati 'Information Landscapes' o 'Paesaggi di Informazione'.

2 Paesaggi di Informazione

Il termine 'Paesaggi di Informazione' deriva dalla dizione anglosassone 'Information Landscapes' coniato da Muriel Cooper, fondatrice del Visible Language Workshop all'interno del MIT Media Laboratory, e descrive un Ambiente Virtuale in cui informazioni testuali vengono rappresentate su superfici comunque disposte nello spazio tridimensionale ed in cui l'utente, o lettore, può 'navigare' pur mantenendo una completa percezione del contesto.

Come caso particolare di Ambiente Virtuale, anche i Paesaggi di Informazione rappresentano un ambiente tridimensionale di sintesi, generato nel calcolatore con il quale un utente è in grado di interagire in modo realistico mediante diverse modalità sensoriali quali la vista, l'udito ed il tatto.

La caratteristica evidente che differenzia un Paesaggio di Informazione da un Ambiente Virtuale convenzionale è costituita dal fatto che in un Paesaggio di Informazione le uniche entità virtuali rappresentate sono le diverse superfici su cui vengono riportate le varie parti di testo con il quale l'utente andrà successivamente ad interagire (SPARACINO 2000). A differenza di un Ambiente Virtuale in cui gli oggetti costituenti possono rappresentare una replica più o meno realistica di oggetti o esseri viventi appartenenti al mondo reale, in un Paesaggio di Informazione gli unici oggetti presenti risultano gruppi di testo scritto disposti opportunamente su superfici dalla geometria la più varia.

David Small, allievo di Muriel Cooper, usa la metafora della costellazione nello spazio per indicare l'effetto visivo generato da uno dei possibili ammassi di testo rappresentati in un Paesaggio di Informazione ed osservato da una naturale distanza all'inizio della procedura di navigazione nell'Ambiente Virtuale che lo rappresenta (SMALL 1994; ID. 1998).

Si immagini infatti di potersi muovere nello spazio 3D e di scorgere all'orizzonte un ammasso chiaro di stelle, appunto una sorta di galassia, che, a mano a mano che avvicina, prende sempre più una forma definita fino al punto di consentire all'utente stesso di poter riconoscere tale ammasso come un insieme di frasi e parole (fig. 1).

Esempi classici sull'argomento sono rappresentati da Information Landscapes costruiti andando a considerare tutte le opere di Shakespeare: in questo caso ogni tragedia può essere posta su un piano nello spazio parallelo ad altri piani che contengano ciascuno il testo di altre tragedie. In un particolare piano il testo della singola tragedia può essere disposto secondo colonne verticali parallele che vanno a rappresentare

i singoli atti. A sua volta la scena di un atto può essere rappresentata come un blocco all'interno della colonna distanziato dagli altri e così



Fig. 1. Spazializzazione di un documento

di seguito. Ne deriva che, nell'ambito di una singola opera, possano essere utilizzate le elevate capacità degli odierni calcolatori per la grafica tridimensionale allo scopo di rendere agevole, sin ad un primo impatto visivo, l'intera struttura del corpo del testo e consentire così all'utente l'individuazione immediata delle parti di proprio interesse. Si ottiene in questo modo una reale interazione visiva con il testo rappresentato aggiungendo in più, rispetto ad una rappresentazione convenzionale di una pagina piena di un libro, la percezione della posizione ed orientazione all'interno del corpo totale del testo.

In un 'Paesaggio di Informazione' l'utente può muoversi liberamente da una superficie all'altra, scegliendo di percorrere traiettorie a propria scelta e quindi 'non vincolate' oppure seguendo percorsi pre-impostati dal motore informatico dell'Ambiente Virtuale stesso. Posizionandosi in corrispondenza di una superficie, l'utente è in grado di prendere visione del testo posizionato nella medesima ed, al contempo, richiamare note o informazioni collegate su altre superfici (KROHN 1995).

Le principali caratteristiche di un 'Paesaggio di Informazione' riguardano:

- Le modalità di presentazione del testo su superfici disposte nello spazio 3D.
- La possibilità di correlare e richiamare parti di testo a seconda dell'interesse dell'utente ed, in alcuni casi, effettuare queste operazioni in modo anche automatico.
- Le modalità ed i processi di interazione tra utente e Paesaggio,

sia per quanto riguarda la navigazione che per il richiamo a contatto della informazione testuale.

Le caratteristiche descritte devono essere progettate, organizzate e finalizzate a garantire un elevato coinvolgimento emotivo dell'utente, una diminuzione dei tempi necessari alla estrazione e ricerca della informazione e, in ultimo, a coinvolgere l'utente in una esperienza piacevole sia in termini estetici che funzionali.

3 *Il Paesaggio come mappa*

L'informazione, strutturata o non, si presta a numerose interpretazioni e visualizzazioni. Nelle prime fasi della progettazione di un Paesaggio Informativo si decide come i testi e le informazioni ad esse associate, i metadati, vengono mappati nello spazio tridimensionale. In funzione del tipo di informazioni si possono utilizzare varie forme, alcune primitive come piani, sfere e cilindri altre più strutturate come griglie, sistemi solari ed alberi. La mappa sta proprio nell'associare una posizione spaziale all'entità informativa, che può essere costituita da un libro o da una pagina dello stesso. In questo contesto si prendono in considerazione Paesaggi Informativi con mappe statiche, nelle quali l'autore ha deciso il modo di strutturare l'informazione, così da ottenere un Paesaggio coerente e curato.

Identificare una buona mappa è importante perché è quella che guida l'utente nell'esplorazione, e che rimane nella sua memoria viva. Tra gli obiettivi della ricerca sui Paesaggi Informativi c'è anche quello di investigare i principi cognitivi che guidano la nostra esplorazione delle informazioni e la valutazione della qualità di una mappa per un certo contesto informativo.

4 *Paesaggi letterari*

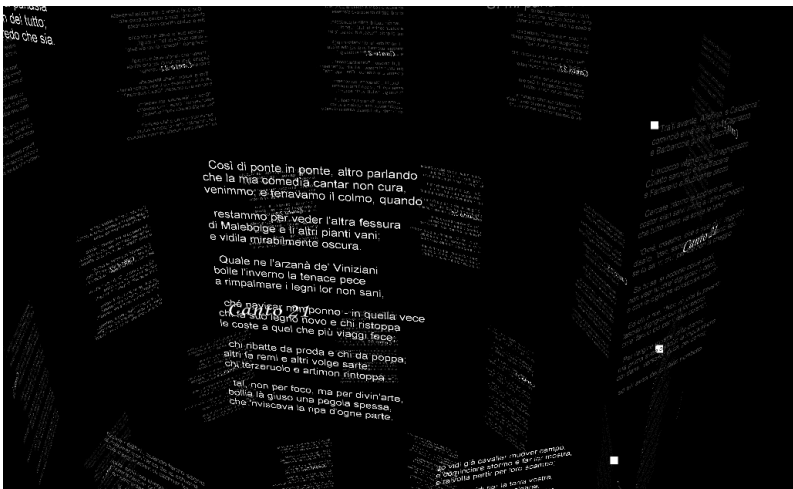
I Paesaggi di Informazione basati su opere letterarie esemplificano il concetto di Information Mapping poiché il contenuto dell'opera viene presentato secondo una metafora visuale associata alla struttura dell'opera o alla sua ambientazione. Le pagine del testo vengono mappate su strutture spaziali diverse dall'ordine sequenziale di un libro, e le parole stesse si trovano ad essere decomposte e ristrutturare. Secondo questa filosofia sono stati elaborati i Paesaggi di Informazione presentati di seguito.

Questi Paesaggi fanno parte dell'attività di ricerca intrapresa negli ultimi anni che ha portato allo sviluppo di uno strumento per la creazione e visualizzazione di Paesaggi di Informazione. Lo strumento di creazione del Paesaggio consente di strutturarlo in aree nelle quali vengono collocati elementi testuali, grafici e tridimensionali, che possono essere messi in relazione attraverso collegamenti ipertestuali. Il posizionamento degli elementi nello spazio avviene in modo simile a quello dei programmi di grafica tridimensionale privilegiando però una organizzazione basata su forme predefinite, in funzione della mappa che è stata scelta.

Ogni Paesaggio ha poi le proprie peculiarità che dipendono dall'opera di riferimento. La disposizione spaziale, i percorsi di navigazione ed in generale le soluzioni estetiche scelte rendono unico ciascun Paesaggio.

Il primo esempio è rappresentato dall'*Inferno* della *Divina Commedia*. I testi dei canti sono stati disposti a ricostruire il girone infernale, creando un duplice effetto sull'utente. La lettura si arricchisce di un contributo emotivo creato dal contesto, e allo stesso tempo l'utente associa i canti alla loro posizione spaziale, stimolando la memoria visiva. Il Paesaggio si apre con una visione globale dell'*Inferno* che colpisce per la vastità dell'opera (fig. 2). In questa fase si apprezza più la complessità e la struttura che non il testo in se. L'utente è poi guidato all'inizio dell'opera di fronte al testo del primo canto, con la possibilità di seguire il flusso logico e temporale dell'opera. Il passaggio ad un'altra parte dell'opera avviene visivamente, navigando giù tra le pagine fino al girone corrispondente (fig. 3). In questo modo in ogni istante il lettore si rende conto in quale parte dell'opera si trovi (fig. 4).

Il secondo esempio è costituito dal Paesaggio ispirato al *Piccolo Principe* di d'Antoine de Saint-Exupéry (fig. 5). La parte del libro inerente ai pianeti è stata resa direttamente nel Paesaggio distribuendo i pianeti nello spazio e raffigurandoli secondo i disegni originali dell'autore. Il testo è stato quindi decomposto su più piani seguendo disposizioni spaziali, sulla base della dinamicità dell'azione del testo stesso. **Q u e s t a** soluzione è ispirata ai calligrammi, immagini di parole, nelle quali la parola non è contenuto ma anche forma. La visualizzazione tridimensionale consente infatti al testo di comunicare in modo diverso, non solo attraverso la relazione spaziale tra le parole ma anche la tangibilità e la forma dei caratteri. Questo è reso possibile da un sistema tipografico virtuale che consente di rappresentare il testo sia su un foglio virtuale dalle forme diverse che come un'entità tangibile. La tangibilità del testo

Fig. 2. Apertura del Paesaggio della *Divina Commedia*Fig. 3. Lettura di un Canto della *Divina Commedia*

e l'uso strutturale dello stesso nello spazio prende ispirazione da quelli che sono i principi della tipografia dimensionale.

L'altro aspetto che emerge dal Paesaggio del *Piccolo Principe* è costituito dai collegamenti tra le parti del testo. Il concetto di navigazione di un ipertesto prevede il passaggio da un argomento all'altro con un cambio di pagina istantaneo, che annulla completamente il contesto visivo dell'argomento di partenza per spostarsi a quello di destinazione. Nel

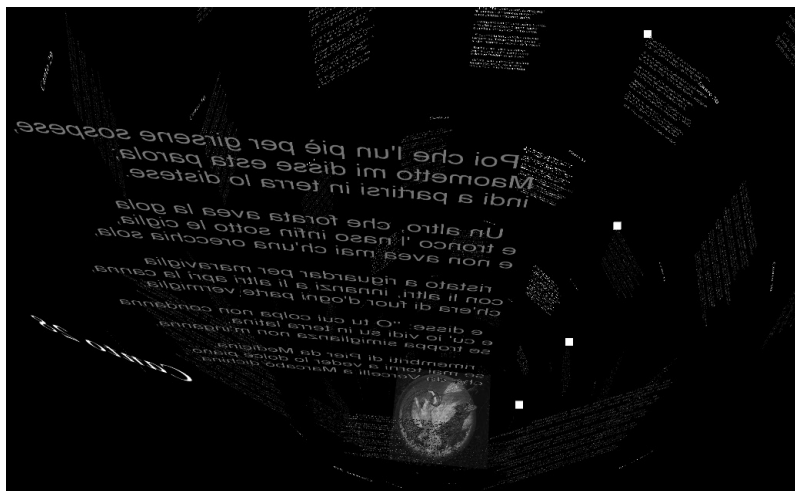


Fig. 4. Navigazione all'interno dei gironi

Paesaggio Informativo, invece, il collegamento ipertestuale corrisponde ad una navigazione da un punto all'altro dello spazio, così da rendere meno forte il distacco visivo dall'informazione iniziale. Il collegamento ipertestuale diventa così un vero e proprio viaggio che può portare l'utente ad esplorare il Paesaggio attraverso un percorso con i suoi punti di osservazione ed i suoi tempi (fig. 6). Questo Paesaggio, rivolto principalmente ad un bambino, invita alla lettura dell'opera coniugando rappresentazioni dei personaggi con una disposizione del testo ritmica e movimentata.

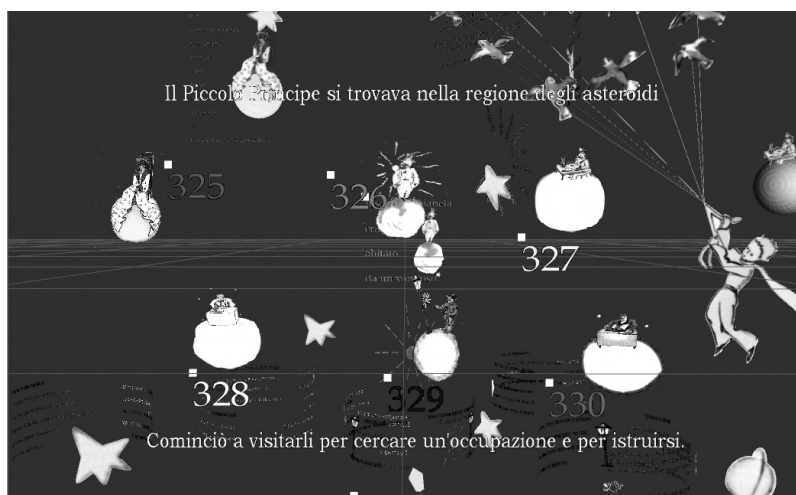


Fig. 5. Gli Asteroidi del Piccolo Principe

5 Visioni future

I principi e le metafore alla base dei Paesaggi Informativi permettono di creare mondi e visualizzazioni affascinanti e ricchi di informazioni. La sfida nella realizzazione di questi Paesaggi consiste proprio nel gestire la complessità e permettere all'utente una fruizione efficace e piacevole. La navigazione in un Paesaggio Informativo complesso, nel quale non vi sono elementi architettonici o paesaggistici tradizionali può indurre nell'utente una sensazione di smarrimento. La prospettiva è quindi quella di valutare quali elementi della navigazione inducano tale smarrimento e di fornire stimoli visivi periferici che rendano l'utente consapevole della sua posizione.

L'altra prospettiva per i Paesaggi Informativi consiste nel diventare più flessibili rispetto al modo di pensare dell'utente che deve poter essere in grado di manipolare la mappa ed interagire con il testo. Dando la possibilità di trasformare la mappa si consente all'utente di esplorare l'informazione in modo nuovo, mettendo in evidenza strutture o elementi non visibili nel Paesaggio originale. Quando poi si attribuisce la possibilità di integrare nuovi testi e di poter collegare quelli esistenti ci si muove verso la sperimentazione di nuove forme di scrittura, che può essere arricchita dai contributi di altri utenti quando il Paesaggio Informativo è di tipo collaborativo.

Dal punto di vista dei contenuti ci stiamo rivolgendo alla creazione di Paesaggi basati sulla vita e le opere di Galileo Galilei. Da un lato sfruttando la metafora della costellazione per fruire il testo originale assieme alla trascrizione e alla sua traduzione, dall'altro evidenziando elementi della vita dello scienziato in cui si mettono in relazione le persone, i luoghi e la sua produzione scientifica.

In sintesi, in una visione più ampia, i Paesaggi Informativi sono a nostra disposizione non solo per esplorare informazioni complesse ma anche per leggere in modo diverso le opere, potendo diventare essi stessi una forma di espressione artistica.

MASSIMO BERGAMASCO, EMANUELE RUFFALDI, CHIARA EVANGELISTA

Bibliografia

D.A. BOWMAN *et al.*, *Information-rich Virtual Environments: Theory, Tools, and Research Agenda*, in *Proceedings of the ACM symposium on Virtual reality software and technology*, New York, ACM Press 2003, pp. 81-90

V. BUSH, *As We May Think*, in *Interactions*, 3, New York, ACM Press

1996 (1945), pp. 35-46

S.K. CARD et al., *The Perspective Wall: Detail and Context Smoothly Integrated*, in *Proceedings of the Conference on Human Factors in Computing Systems CHI'04*, New York, ACM Press 1991

S.K. CARD et al., *3Book: A 3D Electronic Smart Book*, in *Proceedings of the Conference on Human Factors in Computing Systems CHI'04*, New York, ACM Press 2004

C. CHEN, *Information Visualisation and Virtual Environments*, New York, Springer 1999

A. DIEBERGER, *Navigation in Textual Virtual Environments using a City Metaphor*, in *Unpublished doctoral dissertation*, Department for Design and Assessment of Technology, Vienna, Technical University of Vienna 1994

J.A. MILLER, *Dimensional Typography: Case Studies on the Shape of Letters in Virtual Environments*, Princeton, Princeton Architectural Press 1996

O. CHAPUIS, N. ROUSSEL, *Copy-and-Paste Between Overlapping Windows*, in *Proceedings of ACM CHI 2007 Conference on Human Factors and Computing Systems*, New York, ACM Press 2007

E. RUFFALDI, C.A. AVIZZANO, M. BERGAMASCO, *Explorative Information Spaces with the Virtual TeXt-o-Graphic Library (VTEXTGL)*, in «VR International in HCI», Las Vegas 2005

C. RUSSO DOS SANTOS et al., *Mapping Information onto 3D Virtual Worlds*, in «IV», London, 2000, pp. 379-386

D. SMALL, S. ISHIZAKI, M. COOPER, *Typographic Space*, in *Conference on Human Factors in Computing Systems*, New York, ACM Press 1994, pp. 437-438

D. SMALL, *Navigating large bodies of text*, in «IBM Systems Journal», XXXV, 1996, pp. 515-525

F. SPARACINO, G. DAVENPORT, A. PENTLAND, *City of News: Cataloguing the World Wide Web through Virtual Architecture*, in «KOS», July 2000, pp. 179-180

U. KROHN, *Visualization of Navigation Retrieval in Virtual Information Spaces*, in *Workshop on New Paradigms in Information Visualization and Manipulation*, Baltimore 1995

T. WHITE, D. SMALL, *An Interactive Poetic Garden*, in *Conference on Human Factors in Computing Systems*, New York, ACM Press 1998, pp. 335-336

Combinazioni di parole e spazi semantici: un'analisi computazionale dei testi di Giordano Bruno

1 Introduzione

Una delle sfide più ardue nell'analisi dei testi filosofici è riuscire a ricostruire il percorso di variazione semantica di una parola all'interno dell'opera di un autore. In questo ambito, infatti, la stessa natura polisemica di una parola è un dato estremamente significativo e pregnante per l'esegesi testuale, in quanto si connette direttamente al modo in cui il filosofo tratteggia la sua riflessione e dialoga con la tradizione filosofica. Inoltre, una stessa parola può subire forti variazioni semantiche sia a livello 'intratestuale', come conseguenza dello stile e delle modalità dell'argomentazione filosofica, sia sul piano 'intertestuale', come effetto dello sviluppo temporale della riflessione dell'autore. Ne risulta che ogni singolo testo filosofico definisce il proprio *spazio semantico* nel quale le parole si trovano a essere collocate in posizioni differenti a seconda delle somiglianze e associazioni che hanno all'interno del testo, e dei rapporti che intrattengono con altre parole.

Questo lavoro è stato svolto nell'ambito di un progetto di collaborazione con il Dipartimento di Filosofia dell'Università di Pisa e il Centro Signum della Scuola Normale Superiore. A tale proposito desideriamo ringraziare per il contributo offertoci in questa ricerca Simonetta Bassi, Daniele Esposito, Francesca Dell'Omodarme, Fabrizio Meroi, Olivia Catanorchi, Elisa Fantechi. Desideriamo anche ringraziare Gwendolyn Giudetti che ha curato il lavoro di lemmatizzazione degli *Eroici Furori* di Giordano Bruno nell'ambito di uno stage del Corso di Laurea in Informatica Umanistica dell'Università di Pisa. L'articolo è stato congiuntamente scritto dai due autori. Per gli scopi specifici dell'Università italiana, Alessandro Lenci (Dipartimento di Linguistica 'T. Bolelli' Università degli Studi di Pisa Via Santa Maria 36 - 56100 (PISA) alessandro.lenci@ling.unipi.it) è responsabile delle sezioni 1, 2, 4, e 5, mentre Felice Dell'Orletta (Dipartimento di Informatica Università degli Studi di Pisa Largo B. Pontecorvo, 3 - 56100 (PISA) felice.dellorletta@ilc.cnr.it) è responsabile della sezione 3.

La possibilità di ricostruire gli spazi semantici che organizzano il lessico di un autore facendoli emergere direttamente dalla tessitura del testo può dunque offrire una prospettiva interessante e innovativa per l'analisi filosofica. Dall'altro lato, è immediato osservare come metodi ormai consolidati di analisi informatica del testo basati su semplici liste di concordanze non siano sufficienti a costruire una mappa semantica di un testo, né tanto meno ad esplorare la sua evoluzione e la sua natura intrinsecamente multidimensionale. L'obiettivo di questo lavoro è mostrare come nuove prospettive di indagine possano venire dall'applicazione all'analisi dei testi filosofici di metodi avanzati di linguistica computazionale e di rappresentazione del significato basati sulla costruzione automatica di spazi di similarità semantica determinati dalle modalità con cui le parole si distribuiscono e si combinano in un testo. A tale scopo, verranno illustrati i risultati di alcuni esperimenti condotti sugli *Eroici Furori* (1585) di Giordano Bruno, mostrando come le tecniche di semantica computazionale permettano di cogliere aspetti interessanti delle dinamiche del senso che si realizzano nel testo e che possono costituire tracce utili all'esplorazione del pensiero bruniano. Nella prossima sezione discuteremo i presupposti teorici dei metodi di semantica computazionale che costituiscono il quadro linguistico-epistemologico di riferimento per la nostra ricerca; nella sezione 3 presenteremo il particolare metodo di semantica distribuzionale usato per gli esperimenti. Infine, nella sezione 4 illustreremo i risultati di alcuni esperimenti condotti su una selezione di parole estratte dagli *Eroici Furori* e significative per caratterizzare alcune aree concettuali particolarmente salienti nella semantica del lessico di Bruno.

2 Modelli combinatori del significato

In verità la cosa vicina inferiore è spinta dalla stretta somiglianza
alla cosa più vicina superiore attraverso alcuni gradi [...]
Perciò attraverso una comune somiglianza si verifica l'accostamento
dalle ombre alle tracce, dalle tracce alle immagini speculari,
da queste a altre cose.

Giordano Bruno, *De umbris idearum*

Descrivere il contenuto semantico di una parola implica affrontare almeno tre ordini di problemi:

I. distinguere i vari sensi o significati che la stessa parola può assumere nei diversi contesti linguistici;

II. rendere conto delle inferenze che il riconoscimento di uno di questi sensi autorizza;

III. catturare le sue relazioni di similarità con altri termini del lessico.

Tali questioni sono modi diversi del presentarsi di un'unica domanda fondamentale: come individuare e caratterizzare il contenuto concettuale di una parola. Secondo un paradigma di ricerca ben consolidato sia in linguistica che nelle scienze cognitive ed in filosofia, una risposta adeguata a tale problema può venire (soltanto) dall'elaborazione di un modello soddisfacente per la rappresentazione del contenuto concettuale. In questa prospettiva, dunque, la descrizione delle proprietà semantiche dei termini lessicali consiste essenzialmente in un problema di rappresentazione della conoscenza. Come tale, la caratterizzazione dello spazio semantico avviene selezionando un opportuno sistema di simboli concettuali e proiettando su di esso i termini lessicali. L'architettura e la geometria del sistema garantiscono la coerenza e correttezza delle inferenze, mentre il problema della descrizione dei significati dei termini lessicali si trasforma in quello di individuare le corrette modalità della loro proiezione sui simboli concettuali.

Sistemi di simboli concettuali per la rappresentazione della conoscenza sono anche noti come ontologie. Un'ontologia definisce, infatti, l'insieme dei concetti rilevanti per descrivere un dato dominio di conoscenza, insieme alle relazioni e agli assiomi che ne costituiscono l'architettura (GRUBER 1995; GUARINO 1998). Sul piano della rappresentazione della conoscenza lessicale, le ontologie forniscono sistemi di tipi o categorie semantiche per la descrizione del contenuto concettuale. Infatti, sensi diversi dello stesso lemma vengono proiettati su elementi diversi dell'ontologia, mentre l'architettura del sistema di concetti si fa carico delle relazioni inferenziali tra i sensi delle parole. Inoltre, le relazioni di similarità semantica possono essere rappresentate in termini di relazioni topologiche all'interno del sistema di categorie concettuali. L'esempio forse più illustre di ontologia per la rappresentazione della conoscenza lessicale è *WordNet* (Fellbaum 1998), il cui sistema di 'synsets' (insiemi di termini sinonimi) costituisce di fatto un'ontologia di concetti lessicalizzati.

Al di là del problema di determinare la corretta architettura del sistema di simboli concettuali per la caratterizzazione dello spazio semantico, il paradigma rappresentazionale sembra fondarsi su un'assunzione cruciale che ne segna però anche uno dei suoi limiti più sostanziali. Infatti,

la sua ipotesi portante è per ogni parola esista un repertorio di sensi o contenuti concettuali ad essa associati più o meno ben determinati e determinabili, depositati nella mente dei parlanti, che costituiscono appunto le forme del suo significato. Tale conoscenza è ovviamente il risultato di processi di acquisizione, modulazione e cambiamento, ma questi aspetti sono comunque indipendenti dalla possibilità di rappresentare l'informazione semantica, per poi usarla ed applicarla in vari contesti. In altri termini, l'aspetto delle dinamiche alla base del contenuto concettuale è essenzialmente secondario rispetto alla questione della sua rappresentazione. Le ontologie lessicali sono infatti basate sull'assunzione che modellare il significato implichi rappresentare la competenza lessicale come fondamentalmente distinta dal modo in cui le parole sono usate negli effettivi contesti linguistici.

Questa posizione teorica rinvia alla classica opposizione tra competenza e uso, tipica del paradigma generativo in linguistica. Nella teoria sintattica tale opposizione prende generalmente la forma di una dicotomia irriducibile tra la conoscenza del linguaggio (es. la sua grammatica) e come questa viene usata nei concreti contesti comunicativi. A sua volta tale assunzione implica che le descrizioni grammaticali vengano considerate come essenzialmente indipendenti da qualsiasi distribuzione di uso. In altri termini, la rappresentazione della conoscenza del linguaggio può e deve essere astratta da qualsiasi particolare sistema o modalità che ne regola l'uso (TOWNSEND-BEVER 2001, p. 37). In maniera simile, la maggior parte degli approcci simbolici alla rappresentazione del significato assumono una parallela dicotomia tra il 'contenuto semantico' di una parola e il suo 'uso nei contesti linguistici'. Un'ontologia lessicale di simboli concettuali intende infatti essere una rappresentazione del contenuto semantico, indipendentemente dal suo comportamento in contesto. Una delle conseguenze principali di tale paradigma è che le ontologie lessicali hanno come limite proprio la difficoltà di modellare le dinamiche lessicali che si realizzano nel momento in cui i significati si inverano nei concreti contesti testuali. L'ostacolo maggiore per lo sviluppo di ontologie per la rappresentazione semantica risiede infatti nel carattere peculiare della conoscenza lessicale. Questa ha una natura essenzialmente eterogenea ed un'organizzazione solo implicitamente strutturata. Inoltre, la proiezione degli elementi lessicali su qualsiasi sistema discreto di simboli concettuali che costituiscono un'ontologia è resa estremamente complessa dalla polisemia, come tratto onnipervasivo e fondante del lessico. I sensi delle parole appaiono come realtà multidimensionali, dai confini incerti e spesso sottodeterminati, che malamente si prestano a riduzionistiche proiezioni su sistemi concettuali

non sufficientemente adeguati a rappresentarne la complessità strutturale interna e l'intrinseca variabilità. Come sottolineato in Pustejovsky (1995), una rappresentazione lessicale soddisfacente deve essere in grado di rendere conto della natura proteiforme del lessico e delle sue dinamiche. Queste lo rendono simile a un caleidoscopio di sensi, le cui forme mutano a seconda delle interazioni e dei rapporti che si vengono ad instaurare tra i termini lessicali all'interno dei vari contesti linguistici in cui possono ricorrere.

Le ricerche nelle scienze cognitive hanno inoltre dimostrato come le parole differiscano in maniera determinante per quanto riguarda la salienza relativa delle varie dimensioni concettuali che le costituiscono. Ad esempio, mentre i termini di genere naturale (es. *cane*, *cavallo*, ecc.) sono essenzialmente organizzati in maniera tassonomica, una caratterizzazione adeguata dei termini artifattuali (es. *macchina*, *aeroplano*, ecc.) richiede invece una specificazione della loro funzionalità (KEIL 1989). In modo simile, la rappresentazione del contenuto di termini lessicali astratti, dei verbi, degli aggettivi, ecc. richiede l'individuazione di aspetti differenti e peculiari dello spazio semantico. Esiste anche una ricca evidenza empirica che mostra come i giudizi di similarità semantica formulati dai parlanti siano fortemente dipendenti dal contesto e dalla prospettiva rispetto alla quale sono espressi. (cf. BARSALOU 1982; MEDIN *et al.* 1993). Dall'altro lato, le rappresentazioni standard in termini di ontologie di simboli concettuali non sono tipicamente in grado di modellare tali peculiarità del significato lessicale; infatti nelle ontologie lessicali, i significati delle parole sono essenzialmente entità *context-free*, che non mutano a seconda del contesto, né sono determinati o condizionati da questo.

Esiste però un approccio differente al significato delle parole che rovescia diametralmente il rapporto tra parole e contesto. Secondo questo modello, che si basa sulla cosiddetta 'ipotesi distribuzionale del significato', il contenuto informativo di una parola viene assunto come intrinsecamente radicato nei contesti d'uso, dai quali emerge e dai quali viene plasmato. Questo consente di creare un modello diverso di rappresentazione semantica, che approssima meglio l'idea di un lessico *context-sensitive*, sensibile alle variazioni contestuali e capace di adattarsi e mutare forma a seconda delle diverse situazioni d'uso.

Forse il rovesciamento più radicale dell'approccio simbolico alle rappresentazioni semantiche è costituito dall'ipotesi wittgensteiniana che la conoscenza lessicale sia semplicemente un riflesso dell'uso linguistico. Più di recente, questa assunzione è stata adottata da numerosi modelli linguistico-computazionali per la creazione di spazi di similarità semantica basati sulle modalità combinatorie delle parole nei testi. Fin

da Harris (1968) l'informazione distribuzionale ha giocato un ruolo importante per modellare vari aspetti del linguaggio; tale ruolo è ormai ampiamente riconosciuto nella letteratura scientifica anche per lo sviluppo di modelli computazionali del lessico. L'ipotesi distribuzionale viene utilizzata come presupposto teorico di modelli di aspetti diversi del linguaggio, quali il *priming* lessicale (LUND *et al.* 1995), la selezione di termini sinonimi (LANDAUER-DUMAIS 1997), il ragionamento analogico (RAMSCAR-YARLETT 2002) e i giudizi di similarità semantica (MCDONALD-RAMSCAR 2001).

L'ipotesi distribuzionale è di fatto l'erede diretto di una tradizione associazionista e combinatoria che assume come chiave fondamentale per esplorare lo spazio semantico di una parola la ricostruzione delle modalità secondo le quali essa si combina con altre parole. Tale ipotesi trova la sua enunciazione paradigmatica nelle parole del linguista inglese J. R. Firth: «You shall know a word by the company it keeps». Nella sua enunciazione più comune, infatti, l'ipotesi distribuzionale assume che «due parole sono tanto più semanticamente simili, quanto più tendono a ricorrere in contesti linguistici simili». Le modalità con cui le parole si distribuiscono e si combinano tra loro sono dunque come 'tracce' lasciate nei testi dai loro significati. Percorrendo tali tracce, è possibile ricostruire il modo in cui i significati emergono dai loro contesti d'uso e in essi si plasmano. Inoltre, data l'ipotesi distribuzionale, è possibile applicare ai testi metodi computazionali per acquisire dinamicamente le proprietà semantiche attraverso elaborazioni matematiche delle distribuzioni delle parole nei testi stessi. Questo insieme di proprietà semantiche intrinsecamente «dipendenti da contesto» fornisce quindi una caratterizzazione dei possibili sensi di una parola che è direttamente ancorata al testo e che è dunque in grado di far riverberare le sue dinamiche.

3 Vettori di parole e spazi semantici

Secondo il paradigma distribuzionale, è possibile modellare il contenuto di una parola come uno spazio semantico che emerge dalle sue modalità di combinazione con altre parole nei testi. La compresenza di parole in un certo segmento testuale può dunque essere assunta come un dato fenomenologico essenziale per ricostruire le sue proprietà semantiche. Tale ipotesi fornisce quindi una nuova chiave di applicazione di metodi computazionali al testo. Questi possono infatti essere usati per navigare le dimensioni combinatorie delle parole e ricostruire da essi una cartografia degli spazi semantici e delle loro dinamiche. In que-

sta sezione vedremo più in dettaglio un particolare metodo di analisi computazionale distribuzionale applicato a uno dei testi principali di Giordano Bruno, *De Gli Eroici Furori* (1585), con l'obiettivo di esplorare mappe semantiche del lessico bruniano concepite come spazi emergenti dalla distribuzione lessicale nel testo.

Esistono vari modelli computazionali di semantica distribuzionale: *Latent Semantic Analysis* (LSA; LANDAUER-DUMAIS 1997), *Hyperspace Analogue to Language* (HAL; LUND et al. 1995), e più di recente *Correlated Occurrence Analogue to Lexical Semantics* (COALS; ROHDE et al. in stampa). Nella nostra ricerca sul lessico di Giordano Bruno è stato utilizzato il modello *InfoMap* (WIDDOWS 2004), che rappresenta una recente variazione della LSA. Il tratto che accomuna queste diverse istanziazioni dell'ipotesi distribuzionale è l'assunto che quantificare la similarità semantica tra due parole sia equivalente di fatto a valutare la misura in cui si sovrappongono i contesti linguistici in cui esse ricorrono. Questi modelli implementano tale ipotesi attraverso un sistema di analisi computazionale formato da tre componenti fondamentali:

1. una rappresentazione vettoriale delle parole;
2. un metodo di riduzione delle dimensioni dei vettori;
3. una misura della distanza dei vettori, che viene interpretata come una misura della distanza semantica delle parole corrispondenti.

All'interno di questo schema generale, i modelli differiscono per vari aspetti sia rappresentazionali che algoritmici usati per esplorare le modalità combinatorie delle parole nel testo, e spesso legati anche alle diverse finalità teoriche o applicative che fanno da riferimento a ciascun modello.

Come si è detto sopra, alla base dei modelli di semantica distribuzionale risiede l'idea che due parole che tendono a combinarsi con insiemi simili di parole, tendono anche a collocarsi in punti dello spazio semantico più vicini rispetto a quelli occupati da parole che invece si distribuiscono in maniera diversa nel testo. Questa assunzione è tipicamente formalizzata utilizzando una rappresentazione vettoriale di ciascuna parola. In altri termini, una parola è rappresentata come un vettore a n -dimensioni, ciascuna delle quali registra il numero di volte in cui la parola compare in un certo contesto. I modelli differiscono per la nozione di contesto che adottano. Ad esempio in LSA, ogni documento di una collezione rappresenta uno specifico contesto; ciascuna dimensione di un vettore di una parola w_i registra, dunque, il numero di volte in cui la parola ricorre in un certo documento. In InfoMap, invece, i vettori registrano 'co-occorrenze' tra parole di un testo: una parola w_i viene rappresentata come un vettore in cui ciascuna dimensione d_{ij} registra il

numero di volte in cui w_i ricorre all’interno di una finestra di n parole prima e dopo la parola w_j , dove n è un parametro fissato empiricamente. In questo caso, il numero delle dimensioni dei vettori è tipicamente uguale o inferiore al numero di tipi lessicali in un testo. Generalmente, si tratta di un sottoinsieme ottenuto eliminando le cosiddette *stopwords*, ovvero parole appartenenti a classi chiuse come articoli, preposizioni, ausiliari, ecc., che non sono significative per determinare le proprietà semantico-distribuzionali dei termini lessicali.

In InfoMap i vettori vengono costruiti a partire da una matrice di co-occorrenza tra parole come quella riportata sotto:

	<i>dico</i>	<i>diana</i>	<i>degno</i>	<i>intelletto</i>	<i>eroici</i>	<i>bontade</i>
furori	0	1	3	0	4	1
dovenir	0	3	1	0	0	1
possuto	0	0	2	0	0	3
furore	6	2	0	3	0	2
atteone	3	4	0	0	0	0
...						

Fig. 1 – Rappresentazione vettoriale delle parole

Ciascuna riga di questa matrice rappresenta un vettore che codifica la distribuzione combinatoria di una parola (es. *furore*) nel testo. Le colonne della matrice sono le dimensioni del vettore, e registrano il numero di volte in cui la parola compare in una certa finestra di testo a sinistra o a destra di un’altra parola. Nell’esempio in figura 1, il vettore associato a *furore* codifica il fatto che questa parola nel testo ricorre 3 volte a una distanza di n parole a destra o a sinistra dalla parola *intelletto*.

Tipicamente i vettori che codificano le proprietà distribuzionali delle parole hanno un numero di dimensioni molto elevato, generalmente nell’ordine delle migliaia o delle decine di migliaia. Come si è detto sopra, per i vettori costruiti a partire da matrici di co-occorrenza tra parole il numero dimensioni dipende dalla lunghezza del testo e in particolare dal suo vocabolario, ovvero dal numero di parole tipo (nel caso degli *Eroici Furori* le dimensioni dei vettori originari sono le 2000 parole più frequenti nel testo). Un ulteriore problema è dato dal fatto che i vettori sono tipicamente ‘sparsi’, ovvero molte delle dimensioni hanno valore uguale a 0. Questo è legato al fatto che le parole, e in particolare quelle semanticamente più ricche, hanno proprietà combinatorie particolarmente selettive e dunque tendono a ricorrere in prossimità di un sottoinsieme limitato di altre parole. L’alto numero di dimensioni con

valore uguale a 0 può però incidere negativamente sul confronto dei vettori, distorcendo i loro effettivi rapporti di similarità. Allo scopo di evitare questi problemi, tipicamente i metodi distribuzionali applicano un algoritmo di ‘compressione’ dei vettori, che ne riduce le dimensioni, pur preservando la distribuzione statistica dei dati. In particolare, InfoMap adotta il metodo della *Singular Value Decomposition* (SVD) – lo stesso usato da LSA – e riduce i vettori originari a vettori tipicamente di 100 dimensioni. Le dimensioni dei vettori originari rappresentano le coordinate della parola in uno spazio n -dimensionale di co-occorrenza tra le parole nel testo. Dopo l’applicazione della SVD, i vettori rappresentano invece le parole in uno spazio con un numero di dimensioni più ridotto e che possono essere interpretate come dimensioni semantiche ‘nascoste’ e implicite nella distribuzione delle parole del testo, che ne costituisce la manifestazione esteriore.

Ciascuna dimensione di un vettore può essere appunto interpretata come un particolare tratto semantico di una parola. Ovviamente non si tratta di tratti semantici come quelli normalmente usati nella descrizione simbolica (es. + UMANO, – CONCRETO, ecc.), bensì di tratti che esemplificano proprietà semantiche della parola implicite nella sua distribuzione combinatoria. Inoltre, questi tratti hanno valori quantitativi che sono ancorati alle occorrenze delle parole nel testo. La misura in cui un dato tratto caratterizza una certa parola w_i , è proporzionale al suo valore quantitativo nel vettore che codifica w_i . Data l’ipotesi distribuzionale, se due parole w_i e w_j hanno valori simili per le stesse dimensioni d_i e d_j , si può dire che w_i e w_j sono semanticamente simili rispetto a tali dimensioni. Questo implica che per determinare la similarità assoluta tra due parole è necessario comparare i loro due vettori rispetto a tutte le dimensioni che li costituiscono. Intuitivamente, maggiore è il numero di dimensioni nelle quali i due vettori presentano valori simili, maggiore è la similarità semantica tra le due parole. Ciò a sua volta presuppone la definizione di una funzione di distanza tra vettori $d(v_x, v_y)$, che associi un valore scalare a una coppia di vettori v_x e v_y che dipenda dal loro grado di similarità. La similarità tra due vettori viene quindi definita come l’inverso della loro distanza, ovvero come un valore quantitativo compreso tra 0 e 1 in modo tale che 0 corrisponde all’assenza di similarità semantica tra due parole e 1 alla loro perfetta uguaglianza. Esistono molte funzioni di distanza che possono essere usate a tale scopo, anche se la più comune è la *Distanza Euclidea*:

$$|v_x - v_y| = \sqrt{\sum_{i=1}^n (v_{x_i} - v_{y_i})^2}$$

La distanza euclidea tra due vettori riflette l'idea che la loro vicinanza topologica nello spazio vettoriale costituisce un'indicazione della vicinanza nello spazio semantico delle parole che essi rappresentano.

Un altro modo per misurare la vicinanza spaziale tra due vettori – e dunque la loro similarità – è dato dal *coseno* dell'angolo che essi formano:

$$\cos(v_x, v_y) = \frac{v_x \cdot v_y}{\|v_x\| \|v_y\|} = \frac{\sum_{i=1}^n v_{x_i} v_{y_i}}{\sqrt{\sum_{i=1}^n v_{x_i}^2} \sqrt{\sum_{i=1}^n v_{y_i}^2}}$$

Il coseno formato da due vettori nello spazio viene assunto come un coefficiente di correlazione tra le parole. Se due vettori sono infatti geometricamente allineati sulla stessa linea nella stessa direzione, l'angolo tra di loro è 0° , e il coseno è 1 (massima similarità); viceversa, se i due vettori sono indipendenti (ortogonali), il loro angolo è vicino a 90° e il coseno di 90° è uguale a 0 (assenza di similarità). Bisogna anche aggiungere che se i vettori sono normalizzati¹, allora la distanza euclidea e il coseno producono un ordinamento di similarità equivalente: in altri termini, se vogliamo sapere quale tra due parole w_i e w_j sia più semanticamente simile a una terza parola w_k , la distanza euclidea e il coseno ci forniscono la medesima risposta.

Possiamo immaginare i concetti lessicali rappresentati distribuzionalmente come se fossero collocati in quello che Gärdenfors (2000) chiama uno *spazio concettuale*. Questo è definito da un certo insieme di dimensioni che impongono una struttura topologica alla distribuzione dei dati empirici. Secondo il classico esempio di Gärdenfors, la struttura dello 'spazio' dei colori è definito da tre dimensioni: tonalità, saturazione e luminosità. Il significato di ciascun termine di colore può dunque essere descritto attraverso un vettore a tre dimensioni i cui valori collocano il termine in una certa posizione rispetto a ciascuno dei tre assi dello spazio. Secondo questo modello, termini di colore semanticamente simili

¹ Un vettore si dice normalizzato se la sua lunghezza è uguale a 1. La 'lunghezza' di un vettore è uguale alla radice quadrata della somma dei quadrati delle sue dimensioni. Per normalizzare un vettore è sufficiente dividere ciascuna delle sue dimensioni per la lunghezza del vettore.

sono collocati in posizioni vicine dello spazio concettuale. In maniera del tutto parallela, i modelli distribuzionali suggeriscono che i vettori a n -dimensioni che codificano co-occorrenze tra parole definiscono uno spazio semantico-lessicale nel quale queste assumono posizioni che dipendono dalla loro distribuzione nei contesti linguistici. Classi di parole semanticamente simili sono così rappresentate come ‘nuvole’ di punti vicini nello spazio vettoriale multidimensionale.

4 Costruire spazi semantici di parole dagli *Eroici Furori*

Gli *Eroici Furori* di Giordano Bruno sono un esempio prototipico di come lo studio dello spettro semantico di una parola sia reso particolarmente impervio proprio dalla retorica e dallo stile di un autore. Gli *Eroici Furori* sono infatti un dialogo in cui agiscono due o più personaggi che assumono diversi ruoli; inoltre, il dialogo è composto di sonetti e descrizioni di figure allegoriche, alle quali si alternano le spiegazioni del loro significato e discussioni filosofiche sul tema espresso dalle allegorie. Il risultato globale è uno spazio semantico complesso e polimorfo nel quale ogni parola può essere usata nell’arco di poche pagine con un significato di volta in volta poetico, figurato, filosofico, ecc. Vediamo quindi come l’applicazione agli *Eroici Furori* di un modello di semantica distribuzionale come InfoMap possa fornire una chiave di accesso diversa per esplorare le complessità dello spazio lessicale del testo di Bruno.

Tipicamente i modelli computazionali come LSA o InfoMap sono applicati a un testo nella sua natura base di sequenza di forme ortografiche (*tokens*), senza alcun processo di normalizzazione linguistica o analisi morfologica. Secondo questa prospettiva, *furore* e *furori* verrebbero dunque considerati come due unità distinte sia per quanto riguarda la rappresentazione vettoriale, sia per quanto riguarda le dimensioni che occupano nella matrice di co-occorrenza. Questo tipo di approccio presenta l’ovvio vantaggio di non richiedere nessun preventivo trattamento o analisi linguistica del testo che vada oltre la sua semplice ‘tokenizzazione’, ovvero la semplice divisione del testo in unità ortograficamente distinte. Dall’altro lato, il prezzo da pagare per questa scelta è quello di perdere generalizzazioni semantiche importanti. Solitamente, infatti, le diverse forme morfologiche di uno stesso lemma condividono le stesse proprietà semantiche. Dal momento che queste sono estratte su base distribuzionale, sarebbe dunque opportuno che le distribuzioni delle diverse forme confluissero in un unico vettore associato al loro lemma di riferimento, e che i rapporti di similarità semantica venissero ricostruiti

confrontando i vettori dei lemmi e non di forme morfologicamente distinte. Chiaramente questo problema è ancora più urgente in una lingua morfologicamente ricca come l’italiano e nel caso dei verbi, dato l’alto numero di forme distinte in cui un lemma verbale si può trovare a essere realizzato in un testo. Per superare questi limiti, abbiamo deciso di usare una rappresentazione più astratta del testo bruniano. A tale scopo, gli *Eroici Furori* sono stati lemmatizzati semi-automaticamente attraverso *Italian NLP*, un sistema integrato di strumenti per il trattamento automatico dell’italiano contemporaneo. La revisione manuale è stata resa necessaria per correggere i fallimenti prodotti dal sistema, dovuti soprattutto a parole dell’italiano bruniano non presenti nel lessico morfologico impiegato da *Italian NLP*.

Il testo lemmatizzato è stato quindi usato per costruire modelli dello spazio semantico degli *Eroici Furori* attraverso InfoMap. Come abbiamo visto in §. 3, nei modelli distribuzionali una variabile essenziale è rappresentata dall’ampiezza del contesto rilevante per la costruzione delle matrici di co-occorrenza. Questo parametro interagisce fortemente con il tipo di relazioni di similarità tra parole che possono essere ricavate dal testo: una finestra di contesto troppo ampia può infatti portare a individuare associazioni di significato molto ‘lasche’, viceversa una finestra di contesto troppo ridotta può impedire di trovare generalizzazioni semantiche significative o limitarsi ad associare parole semplicemente sulla base di correlazioni idiomatiche. Determinare l’ampiezza giusta della finestra di contesto è un problema empirico e dipende fortemente sia dal tipo di testo che dagli obiettivi di analisi. Nel caso dei modelli elaborati sugli *Eroici Furori*, dopo una serie di esperimenti preliminari, il contesto è stato arbitrariamente fissato in una finestra di 30 parole a sinistra e 30 parole a destra della parola ‘target’, in quanto questa è apparsa soddisfacente per individuare relazioni semantiche interessanti.

La Tabella 1 riporta un frammento dello spazio di similarità costruito con InfoMap per *caccia*, una parola particolarmente significativa per il ruolo che ha nell’opera di Bruno il mito della caccia di Atteone. La prima parola di ogni tabella è la parola target (*caccia* in questo caso), le altre parole sono ordinate per valori di coseno discendente rispetto al vettore della parola target (sono state omesse le parole con un coseno inferiore a 0.5).

caccia	1.000000	caccia	1.000000
<u>cacciatore</u>	0.715892	<u>cacciatore</u>	0.772613
<u>venazione</u>	0.699857	<u>selva</u>	0.640169
<u>cacciatore</u>	0.645397	<u>gemito</u>	0.629727
<u>preda</u>	0.603427	inenarrabile	0.629727
inebriato	0.593269	<u>predare</u>	0.625944
<u>venaggione</u>	0.570792	<u>venazione</u>	0.612536
corporal	0.568281	<u>abbattere</u>	0.612242
gesti	0.564450	<u>sapienza</u>	0.545181
avvegna	0.561344	allargare	0.533222
cattivare	0.560402	penetro	0.528437
vergini	0.558592	comprensibile	0.523042
apelle	0.554898	pelle	0.522736
approvare	0.554297	marcire	0.522024
aggrade	0.541634	intisichire	0.520585

Tabella 1 – parole semanticamente più simili a *caccia* selezionate da InfoMap

La colonna di sinistra della tabella riporta i risultati ottenuti da InfoMap sul testo tokenizzato, mentre la colonna di destra riporta i risultati ottenuti sullo stesso testo lemmatizzato. Analizzando i risultati, possiamo notare come tra i termini a cui InfoMap assegna il valore di coseno più alto con il target si trovino in effetti parole che hanno un rapporto di similarità semantica con *caccia* (es. *cacciatore*, *venazione*, *preda*, ecc.). Il modello vettoriale di InfoMap è così in grado di costruire uno spazio di similarità semantica per le parole nel testo, e il coseno tra i vettori di due parole fornisce dunque una misura indicativa del loro grado di associazione semantica. Detto questo, l'analisi delle parole nella Tabella 1 suggerisce due tipi di riflessioni. La prima riguarda la rappresentazione del testo. La versione lemmatizzata non solo fornisce una quantità maggiore di parole effettivamente associate a *caccia* dal punto di vista semantico, ma presenta anche elementi interessanti come i verbi *abbattere* o *predare*, che sono assenti nella versione ricavata dal testo tokenizzato. Questi emergono come significativamente associati al target esattamente perché sono lemmi che uniscono in sé la distribuzione delle diverse forme attestate negli *Eroici Furori*. Ciò conferma che l'uso di rappresentazioni testuali più astratte e linguisticamente motivate permette un notevole guadagno informativo nell'applicazione di metodi di semantica distribuzionale.

Un altro risultato interessante è la comparsa nella lista delle parole associate a *caccia* di *sapienza*. A prima vista e in astratto queste due parole non sembrano condividere niente dal punto di vista semantico. In

realtà, la loro associazione diventa estremamente rilevante nell'ambito specifico degli *Eroici Furori*, in quanto Bruno usa il mito della caccia di Atteone come allegoria della ricerca della sapienza: «Atteone significa l'intelletto intento alla caccia della divina sapienza, all'apprensione della beltà divina» (*Eroici Furori*, parte I, dialogo IV). Questo ci conduce quindi a una seconda e più sostanziale riflessione sui metodi distribuzionali e che riguarda proprio la conformazione dello spazio semantico individuato da InfoMap. Le parole a cui il modello computazionale assegna i valori più alti di coseno rispetto alla parola target di fatto catturano un ambito vario e diversificato di ciò che possiamo chiamare similarità semantica. Limitandoci al caso del modello ricavato dal testo lemmatizzato, possiamo vedere come tra le parole giudicate più simili a *caccia* si trovino veri e propri sinonimi o quasi-sinonimi – *venazione* –, parole che identificano partecipanti tipici all'evento e che sono anche associate a *caccia* anche sul piano morfologico – *cacciatore* –, nomi che designano il luogo tipico in cui la caccia si svolge – *selva* –, eventi e azioni legati alla caccia – *predare* o *abbattere* – per giungere poi a parole come la già menzionata *sapienza* che hanno un legame metaforico e allegorico con il termine target, valore che è strettamente limitato all'uso di *caccia* nel particolare contesto degli *Eroici Furori*. Questo spettro di parole rappresenta un esempio prototipico della conformazione degli spazi semantici costruiti dai modelli distribuzionali, e in grado di cogliere aspetti, sfumature e sfaccettature diverse del significato di una parola. Questi spazi possono infatti contenere simultaneamente parole legate al target da relazioni semantiche di tipo molto diverso e che vanno dalla sinonimia all'iperonimia, dall'antonimia fino alla metafora e all'allegoria. Gli spazi semantici che emergono dalle distribuzioni delle parole nel testo sono dunque ben lontani dalle geometrie razionali delle strutture ontologiche. Dall'altro lato, esse sembrano però in grado di riprodurre proprio quella multidimensionalità che abbiamo detto essere tipica ed intrinseca del significato lessicale, parte della quale è stabile e trasversale rispetto a molteplici usi di una parola e parte della quale invece nasce in contesti specifici. Un termine come *caccia* è in realtà un nodo di associazioni poliedriche con una molteplicità di concetti che caratterizzano il suo contenuto di significato. Possiamo dunque vedere le liste di parole nella Tabella 1 come uno spaccato dell'«intorno semantico» della parola target, un territorio dai confini fluidi, in grado di cambiare fisionomia e plasmarsi nelle sue relazioni semantiche a seconda dei diversi contesti d'uso. Simili considerazioni valgono anche per le parole riportate in Tabella 2: *verità* e *infinito* (in questo caso solo per il modello costruito sul testo lemmatizzato).

verità	1.000000	infinito	1.000000
<u>apolline</u>	0.622923	<u>infinitamente</u>	0.834759
ingiuria	0.622744	<u>finito</u>	0.826582
<u>veritade</u>	0.618802	<u>infinita</u>	0.738326
solitudine	0.603709	positivo	0.736982
<u>inaccessibile</u>	0.594529	<u>potenza</u>	0.655363
<u>inobiettabile</u>	0.593932	<u>interminato</u>	0.638099
<u>infallibile</u>	0.590822	accio	0.636445
<u>visione</u>	0.589427	rebecchina	0.625724
<u>empedocle</u>	0.576162	<u>finitamente</u>	0.621887
<u>sopranaturale</u>	0.573116	<u>perfettivo</u>	0.617568
empire	0.563613	vano	0.615645
misura	0.563482	affermazione	0.593107
<u>bontà</u>	0.558930	pario	0.592869
deserto	0.553126	circolazione	0.579869

Tabella 2 – parole semanticamente più simili a *verità* e a *infinito* selezionate da InfoMap

Si osservino, ad esempio, nella lista di parole associate a *verità* aggettivi come *infallibile*, *inaccessibile*, *inobiettabile*. Anche in questi casi, gli spazi semantici distribuzionali non sono sistemi geometrici di sensi, ma istantanee che tentano di catturare alcune delle tante possibili trasformazioni sfuggenti ed evanescenti in cui il lessico è in grado di manifestarsi, nella sua natura di eterno Proteo linguistico.

Per esplorare meglio la forma degli spazi semantici estratti dal testo è possibile applicare ai vettori generati da InfoMap tecniche di analisi statistica multivariata dei dati come il *clustering* o lo *scaling multidimensionale* (MDS), che consentono di individuare raggruppamenti di parole semanticamente coese e di valutare le loro reciproche relazioni. Nella nostra analisi degli spazi semantici in Bruno abbiamo usato MDS per analizzare i vettori prodotti da InfoMap per alcuni insiemi di parole estratte dagli *Eroici Furori*. MDS è una tecnica esplorativa dei dati che permette di visualizzare n oggetti in uno spazio a k dimensioni, partendo da informazioni relative alla similarità (o dissimilarità) tra ciascuna coppia di oggetti. MDS colloca nello spazio gli oggetti in maniera tale che le distanze tra di essi riflettano i rapporti di similarità (vicinanza) tra i dati. A partire dall'output generato da InfoMap, abbiamo quindi costruito una matrice simmetrica di distanze tra insiemi di parole, come quella riportata nella figura 2.

	ombra	dio	verità	bene	natura	diana	atteone	caccia	cieco	specchio
ombra	0	1.19041	1.02336	1.36279	1.17934	1.19883	1.36178	1.36975	1.35439	1.29858
dio	1.19041	0	1.148	1.14637	1.10948	1.42823	1.30289	1.47813	1.18138	1.36129
verità	1.02336	1.148	0	1.2855	1.27805	1.15793	1.47752	1.26708	1.31806	1.45071
bene	1.36279	1.14637	1.2855	0	1.08432	1.42155	1.48227	1.46241	1.24584	1.47922
natura	1.17934	1.10948	1.27805	1.08432	0	1.23844	1.35061	1.34199	1.17199	1.41783
diana	1.19883	1.42823	1.15793	1.42155	1.23844	0	1.35507	1.06408	1.41632	1.33005
atteone	1.36178	1.30289	1.47752	1.48227	1.35061	1.35507	0	1.09846	1.51685	1.39582
caccia	1.36975	1.47813	1.26708	1.46241	1.34199	1.06408	1.09846	0	1.39836	1.31619
cieco	1.35439	1.18138	1.31806	1.24584	1.17199	1.41632	1.51685	1.39836	0	1.32282
specchio	1.29858	1.36129	1.45071	1.47922	1.41783	1.33005	1.39582	1.31619	1.32282	0

Fig. 2 – Matrice di distanze euclidee tra parole

Le celle della matrice contengono la distanza euclidea tra ciascuna parola con le altre parole dell’insieme; la distanza euclidea è stata calcolata secondo la formula riportata in §. 3, a partire dai vettori ridotti generati da InfoMap con il metodo della SVD. La figura 3 mostra le stesse parole collocate in uno spazio bidimensionale ottenuto applicando MDS alla matrice di distanze.

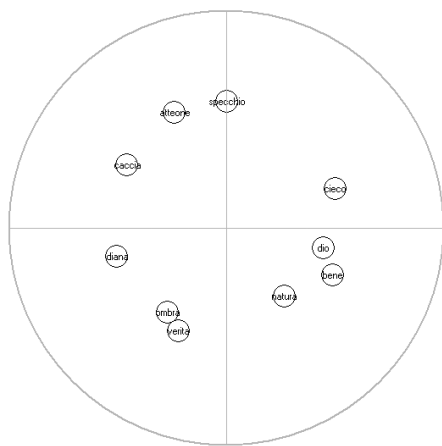


Fig. 3 – Analisi con MDS di un insieme di parole estratto dagli *Eroici Furori*

Si può notare come la rappresentazione spaziale prodotta da MDS rappresenti alcune proprietà interessanti dei rapporti di similarità semantica tra queste parole. Ad esempio, *caccia*, *Atteone* e *specchio* formano un nucleo fortemente ravvicinato e questo trova una corrispondenza nel fatto che tali termini rappresentano protagonisti diversi dell’uso allegorico del mito di Atteone nel testo. Distanza da questo gruppo, troviamo invece un *cluster* altrettanto coeso dal punto di vista semantico in cui compaiono ravvicinati *dio*, *bene* e *natura*. Un elemento interessante

è la posizione che assume nello spazio semantico *Diana*, a metà strada tra il gruppo di parole legate al mito della caccia e la coppia *ombra* e *verità*. Questo è in qualche modo suggestivo del ruolo di Diana nel testo bruniano, appunto come partecipante del mito della caccia e al tempo stesso allegoria della meta a cui giunge Atteone. Come dice Ciliberto, «Atteone in fine vede. Non vede Apollo, vede Diana, non vede la natura produttrice, ma la natura prodotta, non vede la natura generante, vede quella generata. Vede l'universo, l'ombra, il simulacro, lo specchio del 'primo principio'. Ma oltre i numeri e le distinzioni, egli coglie l'unità, gode dell'unità. E qui sta la sua felicità» (CILIBERTO 2004, p. XXXIII).

5 Conclusioni

Le ontologie di simboli concettuali rappresentano architetture formali per la rappresentazione del significato particolarmente utili per catturarne aspetti inferenziali. Dall'altro lato, proprio la natura formale e simbolica di tali sistemi rappresenta un elemento di rigidità che si traduce in un'intrinseca difficoltà per le ontologie ad adattarsi alla multiformità del significato lessicale, alla sua variabilità contestuale e alla sua perenne fluidità. Una buona parte del significato di una parola è infatti costituito da analogie ed associazioni semantiche difficilmente riducibili a inferenze deduttive, molte delle quali vengono create in contesti specifici, spesso tramite relazioni di combinazione e prossimità con altre parole nel testo. Per tali motivi, i modelli distribuzionali meglio si prestano a catturare e a rappresentare tali componenti del significato, che vengono fatti emergere proprio dall'analisi statistica degli spazi combinatori delle parole. Questo risulta ancora più vero quando si pensi all'opera di un autore come Bruno che dell'allegoria e dell'associazione semantica ha fatto una vera e propria *ars inveniendi*.

I modelli computazionali del lessico basati sull'ipotesi distribuzionale permettono di ricostruire e navigare in modo nuovo lo spazio lessicale di un testo. Questi metodi possono dunque essere utili sia a livello confermativo, per convalidare sul testo ipotesi sulla semantica di un autore, sia a livello autenticamente euristico, consentendo di disvelare associazioni e relazioni semantiche spesso nascoste nelle combinazioni delle parole nel testo. Sarebbe ovviamente tanto ingenuo quanto sbagliato pensare che i modelli distribuzionali possano darci un quadro completo della semantica del lessico. Al contrario essi sono strumenti imperfetti i cui risultati vanno costantemente valutati rispetto ai diversi parametri che li possono condizionare: tipo, lunghezza e struttura del testo, il

modo in cui questo viene rappresentato e analizzato a livello linguistico, gli algoritmi per la costruzione dello spazio vettoriale, i metodi di indagine statistica dei risultati, ecc. Gli spazi distribuzionali sono imperfetti, sfumati, 'ombre' di strutture concettuali. La semantica distribuzionale si offre dunque come un complemento più che un sostituto dei metodi simbolici tradizionali. Entrambi ci forniscono prospettive diverse e parziali del significato: ordinata, razionale e statica la seconda, dinamica, magmatica e spesso confusa la prima. Il fatto che tali 'viste' appaiano spesso lontane e opposte fino al punto di sembrare inconciliabili è indubbiamente ancora un segno dei limiti della nostra conoscenza sul linguaggio e sulla cognizione, che ci impediscono di comporre questi pezzi in un panorama complessivo sul lessico. Per tale motivo, le parole di Bruno bene si attagliano a catturare la nostra posizione attuale nello studio del significato: «Là è oggetto finale, ultimo e perfettissimo, non già in questo stato dove non possemo veder Dio se non come in ombra e specchio; e però non ne può esser oggetto se non in qualche similitudine; [...] Or di queste specie e similitudini si pasce l'intelletto umano da questo mondo inferiore, sin tanto che non gli sia lecito de mirar con più puri occhi la bellezza della divinitade» (*Eroici Furori*, parte I, dialogo III).

ALESSANDRO LENCI, FELICE DELL'ORLETTA

Bibliografia

- BARSALOU 1982 = L.W. BARSALOU, *Context-Independent and Context Dependent Information in Concepts*, in «Memory and Cognition», X, 1982, pp. 82-93
- CILIBERTO 2004 = M. CILIBERTO, *Introduzione* a G. BRUNO, *Eroici Furori*, Bari, Laterza 2004, pp. VII-LXI
- FELLBAUM 1998 = *WordNet. An Electronic Lexical Database*, ed by C. Fellbaum, Cambridge, The MIT Press 1998
- GÄRDENFORS 2000 = P. GÄRDENFORS, *Conceptual Spaces. The Geometry of Thought*, Cambridge, The MIT Press 2000
- GRUBER 1995 = T.R. GRUBER, *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, in «International Journal of Human and Computer Studies», XLIII, 1995, 5/6, pp. 907-928
- GUARINO 1998 = N. GUARINO, *Some Ontological Principles for Designing Upper Level Lexical Resources*, in *Proceedings of the First International Conference on Language Resources and Evaluation*, Granada 1998, pp. 527-534
- HARRIS 1968 = Z.S. HARRIS, *Mathematical Structures of Language*, New

- York, Wiley 1968
- KEIL 1989 = F.C. KEIL, *Concepts, Kinds and Cognitive Development*, Cambridge, The MIT Press 1989
- LANDAUER-DUMAIS 1997 = T.K. LANDAUER-S.T. DUMAIS, *A Solution to Plato's Problem: the Latent Semantic Analysis Theory of Acquisition, induction and Representation of Knowledge*, in «Psychological Review», CIV, 1997, 2, pp. 211-240
- LUND et al. 1995 = K. LUND-C. BURGESS-R.A. ATCHLEY, *Semantic and Associative Priming in High-Dimensional Semantic Space*, in *Proceedings of the Cognitive Science Society*, Hillsdale, N.J., Erlbaum Publishers 1995, pp. 660-665
- MACDONALD-RAMSCAR 2001 = S. MACDONALD-M.J.A. RAMSCAR, *Testing the Distributional Hypothesis: the Influence of Context on Judgements of Semantic Similarity*, in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh 2001
- MEDIN et. al. 1993 = D.L. MEDIN-R.L. GOLDSTONE-D. GENTNER, *Respects for Similarity*, «Psychological Review», C, 1993, pp. 254-278
- PUSTEJOVSKY 1995 = J. PUSTEJOVSKY, *The Generative Lexicon*, Cambridge, The MIT Press 1995
- RAMSCAR- YARLETT 2003 = M. RAMSCAR-D. YARLETT, *Semantic Grounding in Models of Analogy: an Environmental Approach*, «Cognitive Science», XXVII, 2003, pp. 41-71
- ROHDE et al. (in corso di stampa) = D.L.T. ROHDE-L. GONNERMAN-D.C. PLAUT, *An Improved Model of Semantic Similarity Based on Lexical co-Occurrence*, in «Cognitive Science», in corso di stampa
- TOWNSEND-BEVER 2001 = D.J. TOWNSEND-T.G. BEVER, *Sentence Comprehension: The Integration of Habits and Rules*, Cambridge, The MIT Press 2001
- WIDDOWS 2004 = D. WIDDOWS, *The Geometry of Meaning*, Stanford, Stanford CA, CSLI 2004

Progettare oggi: dalle forme meccanicistiche alle forme digitali alle forme geneticamente mutate, nella continua ricerca della bellezza

Gli architetti e i designers non amano parlare della bellezza. Pensano di riuscire a costruirla, a produrla e a realizzarla. Cercherò di mostrare come nel mondo del design, della decorazione, del giocattolo, dell'interior design dell'architettura, dell'arte, nell'epoca della sua riproducibilità tecnica, si sono consolidate più idee di bellezza che hanno guidato le diverse pratiche di progetto. L'espressione 'idea di bellezza' è troppo platonica e quindi contraddittoria con la mia convinzione che tende a privilegiare una continua ibridazione di diverse pratiche alla ricerca dell'estetico, ma nella nostra vita di tutti i giorni, e, ossessivamente, nella pratica progettuale, ognuno di noi, consciamente o no, ad un'idea di bellezza sempre si riferisce nelle sue scelte.

Intorno a questo bisogno tutto umano di continuamente sapere ciò che è, o meglio sembra, bello e ciò che non lo sembra, cerco di proporre alcune riflessioni. Cercheremo insieme di mettere a fuoco in opere diverse le 'idee di bellezza' oggi dominanti. Vorrei anche mostrare guardando diversissime opere come l'idea di bellezza, acquisita come dominante, sia comunque sempre parziale, di parte, strumento fondamentale di consenso utile sia per eleggere miss universo che per imporre il nuovo modello di merce, e sempre coesistono in un certo periodo storico molte diverse idee alternative e oppostive di bellezza. L'operazione di pro-

Ho proiettato e commentato le immagini di questo seminario agli studenti della Scuola Normale Superiore di Pisa, a quelli del mio corso di progettazione alla specialistica di product design dell'ISIA di Firenze, all'ESAD (Escola superior des Des Artes e Design) di Matosinhos in Portogallo, agli ingegneri dell'IFTS, corso di design coordinato da Pont-Tech per la Fondazione Piaggio e per l'ISIA. È un insieme di immagini con un inizio di commento assolutamente incompleto. La incompiutezza del testo rispetto alla complessità dei temi trattati è un invito a continuare la discussione (p.degane@paolodeganello.it).

getto è questa ricerca infinita della bellezza che non può dare risultati definitivi, ma a cui sempre a noi progettisti si chiede di tendere, anche imponendoci di dare bellezza alla mostruosità dell'umano, occultandone la sua bruttezza.

Vedremo nella parte finale di questo seminario come oggi, nel mondo dell'arte, è sempre più presente il tema della 'bruttezza' piuttosto che il tema della 'bellezza'.

La bellezza è sempre stato strumento di potere.

Il gruppo dominante impone ai suoi subalterni, ai suoi soggetti dominati un'idea di bellezza a cui uniformarsi; ed esistono sempre dei soggetti sociali, dei gruppi più o meno organizzati o dei singoli che chiamiamo progettisti che cercano di contrapporre all'idea dominante un'altra nuova sorprendente bellezza.

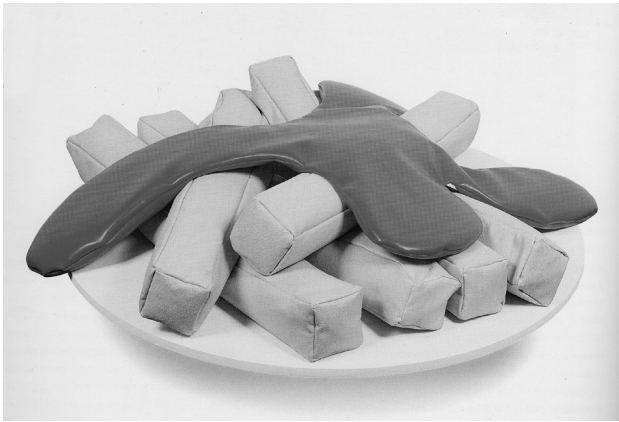
«Continuiamo a scorgere forme impreviste di bellezza che brillano in nuove costellazioni, anche se non sempre vengono immediatamente notate» (R. BODEI, *Le forme del bello*, Bologna, il Mulino 1995). Un progettista non può mai rinunciare a questa fondamentale speranza di «scorgere forme impreviste di bellezza» che rilancino e rilegittimino il suo lavoro. Mi ha molto e positivamente colpito quell'orrendo pullover a striscie colorate che il primo indio eletto presidente della repubblica boliviana, Evo Morales, indossava il giorno della sua vittoria elettorale. Invece di un elegante vestito grigio con cravatta il presidente del popolo del sole si proponeva alle televisioni di tutto il mondo, con un pullover a striscie colorate. Una forma imprevista di bellezza, il riemergere dei segni di una civiltà pre-colombiana, distrutta ed emarginata per secoli, una sorpresa venuta dal sud del mondo, una contromossa stupefacente degli emarginati, nell'epoca della guerra in Iraq può essere il significato da dare a quel pullover a striscie colorate. A me sembra utile per noi progettisti imparare a vedere l'idea di bellezza che emerge dalle contromosse degli esclusi, dei solitari e perdenti, ripeto, nell'epoca della guerra in Iraq. In queste contromosse colloco la mia speranza di bellezza, domandatevi su quale altra speranza di bellezza possiamo contare...su quella imposta dalla cultura dominante che continuamente ci costringe a svolgere lo sguardo da un'altra parte per non vedere le mostruosità, da Auschwitz ad Abu Graib, su cui è perpetuato il suo dominio? [figg. 1, 2].

Questo artista, uno scultore pop, è Claes Oldenburg e questo *Senza titolo* (1966) è invece un'opera di un artista minimalista, Donald Judd, entrambi americani, entrambi venuti alla ribalta nella metà degli anni Sessanta. Partiamo dall'arte per vedere come il mondo dell'arte, sempre, dall'epoca della sua riproducibilità tecnica, è anticipatore in termini



1. Foto di Oldenburg con 'Tube'
2. Donald Judd, Senza titolo, 1966

sperimentali della forma che poi prenderanno le merci, contribuendo alla formazione di un immaginario collettivo che condiziona, ispirerà e orienterà il consenso a moda, arredo, grafica, decorazione, design, interior design, architettura e stili di vita. Questo signor Oldenburg, americano di origini svedesi, gira per la città di New York con in braccio una sua opera: un banalissimo tubetto di dentifricio, dieci volte più grande del reale, fatto in sky, un materiale sintetico lucido colorato, morbido con cui, negli anni '60, si facevano gli impermeabili a basso costo che si compravano nei supermercati. Questo dentifricio se lo porta in collo come un bambino troppo cresciuto, c'è una esibita tenerezza in questo coccolare una banale, seppur gigante, merce da supermercato. Certo Duchamp ha portato nelle gallerie, orinatoi, ruote di bicicletta, grucce e porta bottiglie, ma è un ready made che usa l'oggetto decontestualizzato dentro il mondo dell'arte non lo 'rappresenta', Oldenburg ha rappresentato, ricostruiti con lo sky, tubetti di dentifricio, water con sciaquone, un nero telefono a gettoni, [fig. 3] patate fritte con rosso

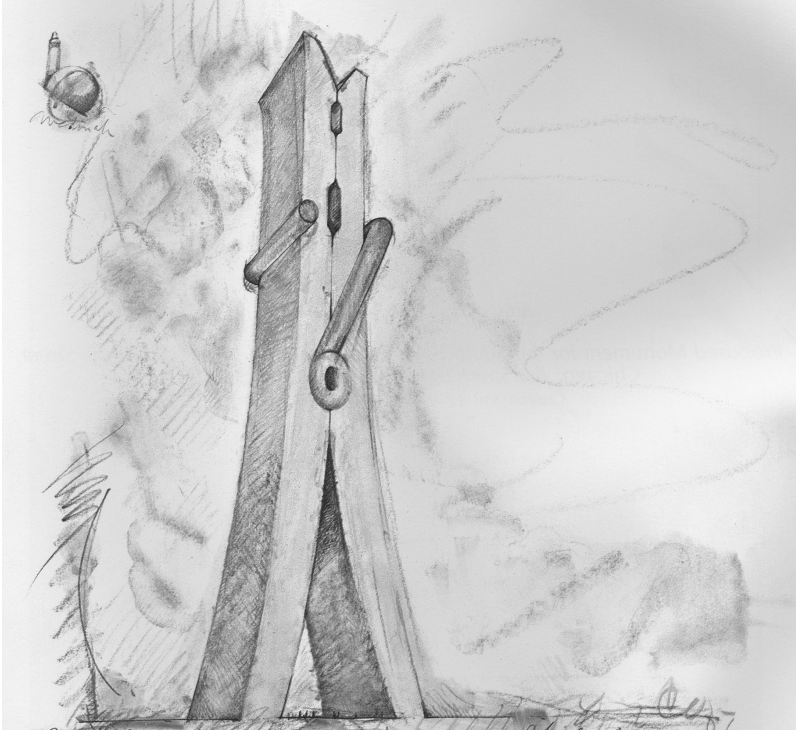


3. Oldenburg, *Patate fritte con rosso kechup*

ketchup. Sono rappresentazioni di oggetti della domesticità, della vita qualunque, immagini di merci private della loro matericità, sono forme flaccide che si sfanno e che danno la percezione del disfacimento, della abbondanza malata della società del benessere americano [fig. 4].

Dal '65 al '69 ha anche progettato una serie di monumenti urbani, quali una molletta per stendere i panni alta 15m, una forbice, alta almeno 50 metri, al posto dell'Obelisco di Washington D.C. [figg. 5, 6].

Ha proposto di spiacciare un gelato pinguino gigante fino aappare la Park Avenue a New York, ha infine realizzato nella piazza Cadorna di Milano, disegnata dalla Gae Aulenti, un ago alto 15 metri inficcato



4. Oldenburg, *Late Submission to the Chicago Tribune architectural Competition of 1922 Clothespin (Version two)*, 1967
5. Oldenburg, *Proposed colossal Monument for Park Avenue, NYC Good Humor Bar*, 1965
6. Oldenburg, *Ago filo e nodo*

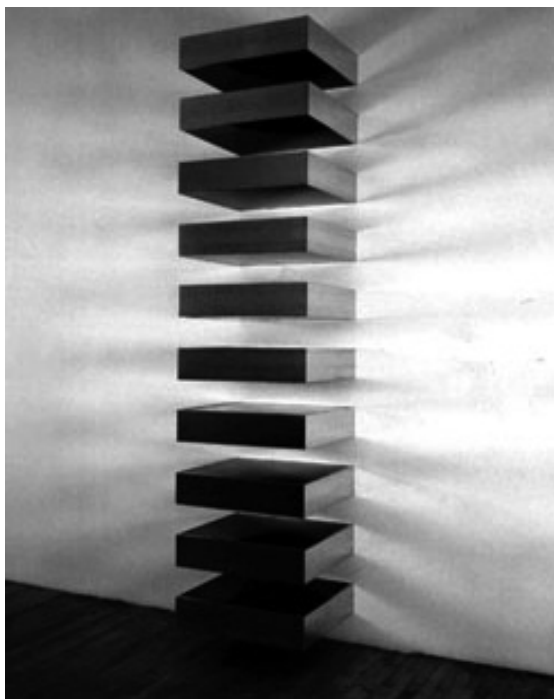
nel terreno, completo di enormi fili colorati, bianchi, rossi e verdi quale monumento alla moda italica.

L'architettura è sempre stata l'elemento fondatore delle arti, ha accolto e posizionato, dato senso alla scultura, alla pittura, alla decorazione, facendola diventare parte del suo disegno, pensate la piazza del Campidoglio disegnata da Michelangelo con al centro il monumento equestre del Marc'Aurelio, provate a mettere una molletta al posto del Marc'Aurelio!

Oldenburg e tutta la Pop Art ci dicono che l'arte comunica attraverso la forma delle merci e attraverso i mass media cioè quegli strumenti di comunicazione che più hanno promosso il consenso sulla merce. La Pop Art ha frantumato i rigidi confini tra arte e merce che vigevano ancora al tempo della Bauhaus. L'arte la si studiava per dare qualità estetica alle merci di grande serie, ma allora l'arte non si interessava delle merci, cercava di darsi un nuovo statuto con l'arte non oggettiva, cioè si rifiutava di rappresentare oggetti cioè realtà verificabili, diceva il pittore Malevic «l'arte non è più interessata a servire lo stato e la religione, non desidera più illustrare la storia dei costumi, non vuole avere più nulla a che fare con l'oggetto come tale e crede nella possibilità di esistere in e per se stessa, a prescindere dalle cose». Il Movimento Moderno che si andava costruendo in quegli anni intorno alla Bauhaus, tutto impegnato a massificare l'estetico elaborato dall'arte attraverso il disegno dei prodotti di serie, non poteva immaginare che quarantanni dopo l'arte si sarebbe messa a rappresentare un piatto di patatine fritte con il ketchup sopra, fatto di sky, o una molletta, o con Warhol la lattina della zuppa Campbel.

Chi disegna le merci, l'arte o il design? [fig. 7].

Contemporaneamente, in assonanza con l'arte non oggettiva di Malevic, Donald Judd proponeva nelle sue opere forme assolutamente geometriche, astratte e fatte di lastre di acciaio, un materiale che permette l'elaborazione di forme perfette, profondamente diverso dallo sky. Quest'opera è composta dalla ripetizione di sei grandi lastre quadrate in acciaio inossidabile, lucido, specchiante, staccate dal muro da delle lastre di plexiglas ambrato che ricostruiscono dei cubi con la facciata in acciaio. L'acciaio specchiato è un materiale che permette all'utente di percepire la sua presenza nell'opera, registra le leggere variazioni di luce e colori dell'ambiente in cui queste opere sono esposte, ma sono forme dure, ripetitive, perfette, eterne come l'acciaio, sono delle astrazioni appena scalfite dalla presenza dell'umano. Lo stesso tema della ripetizione ritmica, della perfezione geometrica delle forme, si ripropone in quest'altra scultura sempre di Judd del 1969 *untitled*. «Togliere



7. Donald Judd, *Senza titolo*, 1969

tutto tranne l'ordine», questo è il commento con cui Angela Vettese in: «Capire l'arte contemporanea» (ALLEMANDI 1998) presenta quest'opera. Togliere tutto tranne un flebile anelito alla variazione; ma non uscire mai dal rigore dell'astrazione geometrica è uno dei principi fondamentali del gusto minimalista. Gli artisti minimalisti, dal loro padre spirituale Franc Stella che aveva presentato nel '59-'60 enormi quadri neri solcati da linee bianche disposte in modo geometrico, a Dan Flavin con le sue composizioni di tubi fluorescenti, a Sol Le Wit autore di grandi grigliati tridimensionali fatti di quadrati che contengono quadrati, costruirono un testo figurativo che è diventato lo 'stile minimalista' oggi dominante nell'arredo, nell'interior design, nella moda, ed è sinonimo di eleganza e di buon gusto. Come sostiene giustamente Fulvio Carmagnola nel suo testo *La conoscenza della conoscenza* sul minimalismo, «l'oggetto minimalista, inabitabile quanto il suo antenato funzionalista, – già Adorno notava quanto poco funzionale fosse il funzionalismo – sembra rendere inabitabile il suo ambiente, generando intorno a se uno spazio rarefatto di assenza, niente simboli se non il simbolo di se stesso. L'oggetto diventa iconoclasta» [fig. 8].

Non è certamente iconoclasta ma tutto pop questo divano disegna-

81031 Aplanas/Pistola/Italy
telefono 71351/2/3

the Imperial Sofa

Composable sofa for the most varied solutions.
Frame in fibreglass, padding in foam rubber,
covering in artificial leopard skin, fabric or leather.
By request, small carpet with supports in expanded resin.

**An imperial piece
in the dreariness of your
domestic walls.
A piece that is finer than you.
A very beautiful piece
that you would not deserve.
Clear your sitting-room!
Clear also your life!**

poltronova

Safari design: Archizoom

il Divano Imperiale

Divano componibile per le più svariate soluzioni.
Struttura in fibreglass, imbottitura in gommapiuma,
rivestimento in finto leopardo, tessuto o pelle.
A richiesta, tappetino con supporti in resina espansa.

**Un pezzo imperiale
nello squallore delle vostre
pareti domestiche.
Un pezzo più bello di voi.
Un pezzo bellissimo
che non meritereste.
Sgombrate il vostro salotto!
Sgombrate anche la vostra vita!**



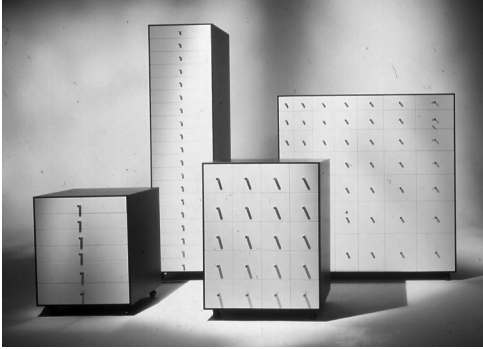
8. Divano Safari di Archizoom Associati, 1969

to nel '69 dal gruppo radicale Archizoom Associati¹, di Firenze per Poltronova. È un cubo di fibreglass bianco o nero ottenuto dall'assemblaggio di due stampi, quello per il singolo posto e quello per il due posti. Sono componenti simmetrici e quindi infinitamente componibili, l'interno del cubo è scavato con una zampata di elefante che ne ricava sei sedili circolari rivestiti di finta pelle di leopardo. L'oggetto è gigante, violenta qualsiasi soggiorno della casa funzionalista, ha un rivestimento, la pelliccia finta di leopardo, che rimanda al vestiario delle avvenenti

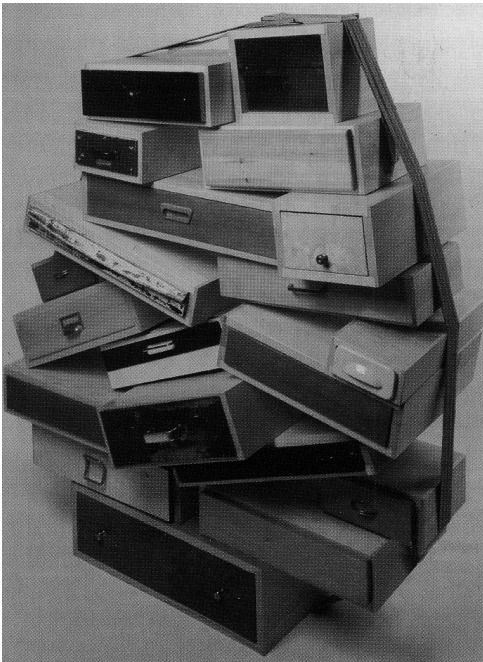
¹ Gruppo dell'architettura radicale attivo a Firenze dal '66 al '73 formato da Andrea Branzi, Gilberto Corretti, Paolo Deganello, Massimo Morozzi, Dario e Lucia Bartolini. Sul gruppo Archizoom Associati cfr. R. GARGIANI, *Archizoom Associati 1966-1974*, Milano, Electa 2007.

donne di colore che col *blakpower* salgono sulla scena americana, è una tipologia di arredo che rivoluziona il tradizionale salotto fatto di un divano tre posti e due poltrone. Col progetto radicale che si alimenta di tutta la cultura del dissenso che accompagna la pop art: la beat generation, gli hippies, il «Do! it! fallo» di Yerry Rubin e le Pantere nere, si sviluppa nella cultura del progetto una diversa attenzione sul disegno dei prodotti che potrebbe essere sintetizzata in uno slogan «reinventiamo la merce» cioè disegniamo un divano oltre ogni opportunismo economico che permetta un nuovo modo di stare insieme, in definitiva, di abitare. L'abitabilità è di più del funzionalismo, dell'utensile divano, proprio perché va oltre la funzione, senza ignorarla ma inglobandola in un'articolazione più complessa della forma, che comunica a diversi livelli: richiama la dissacrante cultura figurativa del negro di America, prende posizione a favore della figuratività dell'oggetto contro la elitaria astrazione, viene incontro ad una diversa domanda dello stare insieme delle nuove generazioni anticonformiste, non più donne impettite sulla sedia o poltrona ma tutti insieme stravaccati a strusciarsi e coccolarsi insieme; e la funzione permane nel morbido del rivestimento peloso. Donald Judd, come dice Angela Vettese, «avrebbe tolto tutto tranne l'ordine», avrebbe lasciato il disegno geometrico e tolta la pelliccia in finto leopardo, riducendo a tal punto quella complessità figurativa che segna la diversità di questo prodotto e in nome della semplicità riducendo al silenzio un prodotto che invece vuole gridare.

È certamente iconoclasta la collezione di cassettiere disegnate da Kuramata per Cappellini [fig. 9]. Una sottile cornice nera racchiude superfici diverse per forma ma sempre quadrate o rettangolari, bianco grigio ulteriormente ripartite in altri rettangoli o quadrati che segnano appena i cassetti il cui centro è marcato da un piccolo cilindro in acciaio satinato con funzione di maniglia. I cassetti della grande cassettera quadrata hanno una particolarità che molto ricorda le costruzioni spaziali di Sol Le Wit, il cassetto più alto a sinistra è molto più piccolo dell'ultimo sulla destra. Questa impercettibile variazione all'interno di una geometria data è l'equivalente della leggera variazione tonale imposta dalla luce alle lamiere di acciaio di Donald Judd, è parte fondamentale del testo figurativo minimalista la variazione minima all'interno delle ferree regole della geometria; provate a confrontare questa variazione minima con l'imprevedibile presenza di pelliccia di leopardo finta nel divano Safari e vi risulterà subito chiaro come il minimalismo cerca di allontanare l'utente dall'oggetto; non si pone il problema dell'abitabilità dell'oggetto, ma solo del virtuosismo del suo disegno. L'oggetto iconoclasta fa il vuoto intorno a sé, considera ogni orma, impronta del vivere,



9. Cassettiere di Shiro Kuramata



10. Cassettiera di Tejo Remy per Droog Design

ogni figurazione, una profanazione della sacralità della forma. Solo una leggera deformazione della geometria per rendere prensile un pomello cilindrico, che è altrimenti scivoloso e faticoso nel gesto di aprire, è ammessa; questa bellezza pulita e silenziosa che il rigore geometrico riesce a dare, sempre ritorna nella ricerca della forma e impone la difficile compatibilità tra abitabilità e progetto. Lo spiritualismo di impostazione platonica degli iconoclasti tendeva ad aborreire dalla materia, ignobile e morta di fronte alla immortale 'realtà' irrapresentabile del Dio e dei suoi Santi. La bellezza per gli iconoclasti e i minimalisti, proprio per la sua dimensione spirituale, è rappresentabile solo attraverso l'astrazio-

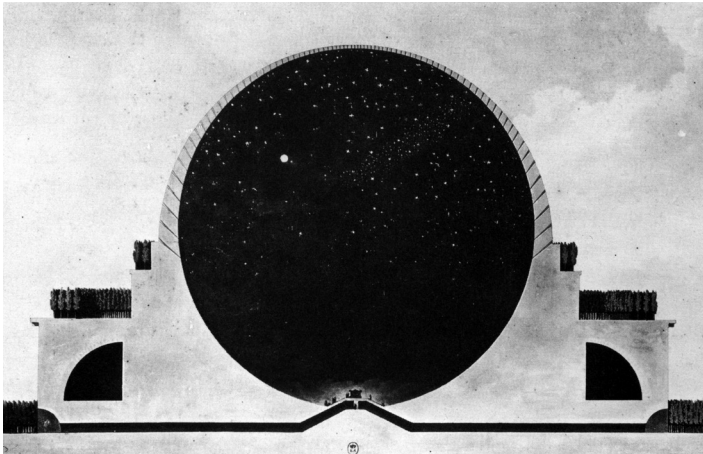
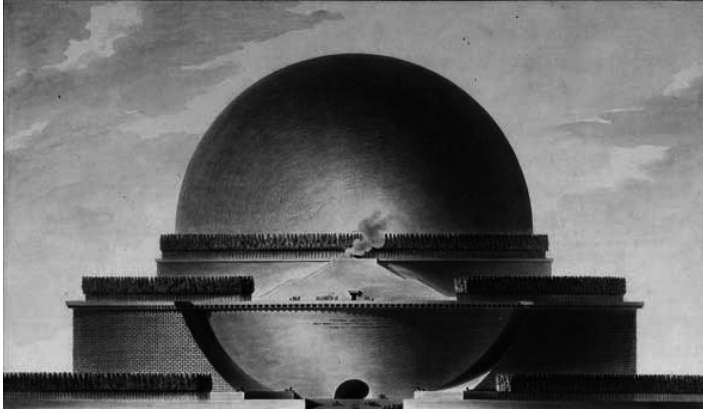
ne, ogni figurazione è peccato. La cassettiera di Kuramata deve perdere, dissolvere la sua figurazione oggettuale, deve liberarsi di tutti i segni e simboli che la riducono a cassettiera e dissolvere la sua identità di utensile nell'astrazione di una superficie rettangolare o quadrata, bianca, incorniciata da un rigo nero trapuntata di pomelli cilindrici di acciaio. Ritorna l'approssimazione al nulla del quadro bianco di Malevic. Qual'è l'astrazione più compiuta dell'umano se non la geometria?

Confrontate le cassettiere di Kuramata con un'opera di un giovane designer olandese del gruppo Drog Design, Tejo Remy [fig. 10], dal complesso titolo «you can't lay down your memory chest of drawers»: una catasta di cassetti di diverse epoche tenuta insieme da una cinghia. Quale tra queste due diverse ed opposte fascinazioni più vi appartiene?

Quando nel 1785 l'architetto Etienne Louis Boullè progettò un Cenotafio per Isac Newton [figg. 11-12a-b], disegnò una vasta sfera in muratura in cui di notte era sospeso un fuoco a rappresentare il sole, mentre di giorno questo veniva spento per svelare l'illusione del firmamento prodotta dalla luce diurna attraverso i muri della sfera perforati con fori a forma di stelle. Frampton dice dell'architettura di Boullè che «l'immensità della visione e la disadorna purezza geometrica della forma monumentale sono combinate in modo da suscitare emozione ed ansia» e che «più di ogni altro architetto illuminista, Boullè fu ossessionato dalla capacità della luce di evocare la presenza del divino»². Già Galileo nel Seicento aveva sostenuto che l'universo «non si può intendere se prima non si impara ad intendere la lingua matematica, e i caratteri son triangoli, cerchi ed altre figure geometriche, senza i quali mezzi è impossibile intendere umanamente parola» e Newton con la legge di gravitazione universale aveva completato quella lingua matematica che permetterà all'astronomo Laplace di rappresentare il cosmo come una gigantesca macchina governata dalle leggi matematiche di Galileo e Newton. La scienza aveva dato la sua rappresentazione dell'universo che sembrava dimostrabile, attraverso il metodo sperimentale, e definitiva.

La perfezione dell'universo svelava quella grandezza di un'opera che solo un artefice divino poteva aver realizzato. La lingua matematica diventava con l'illuminismo la lingua dell'architettura. L'artificiale costruito dall'umano non poteva prescindere dal modello ideale costruito dal divino. La sfera ben rappresentava l'opera di quello scienziato che aveva portato a compimento la conoscenza del creato divino, l'enorme sfera

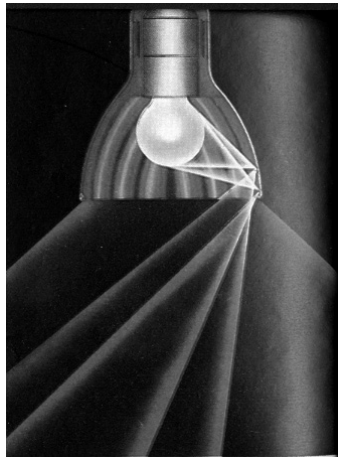
² K. FRAMPTON, *Storia dell'architettura moderna*, Milano, Zanichelli 1993.



11. Boullée, Cenotafio di Newton

12a-b. Boullée, Cenotafio sezione diurna e notturna

13. Meccano Magazine, lampada
Naska Loris



di Boullè era la meritata rappresentazione della grandezza di Newton, la luce divina la illuminava nel suo interno ricordando la piccolezza dell'uomo rispetto alla sublime perfezione del creato. La geometria e il linguaggio della macchina, governata dalle leggi matematiche, costruivano insieme quella figurazione macchinista che diventerà, liberata di tutti gli orpelli di una cultura preilluminista, la forma bella. Tutta la ricerca del movimento moderno fortemente condizionata dallo scientismo tenderà a una continua coniugazione del rigore geometrico con la tecnica e alla progressiva affermazione della oggettiva bellezza della macchina [fig. 13].

La Naska Loris, con le sue molle, i suoi giunti esibiti, il suo miracoloso movimento ben rappresenta questa nuova concezione del bello che coniuga la geometria con il linguaggio macchinista, mostra nella riflessione della luce la necessità geometrica quale rispetto delle regole illuminotecniche ed esibisce le componenti costruttive quali molle giunti

ed aste come una nuova spregiudicata bellezza del moderno.

La nostra generazione si è formata col meccano, un gioco che valorizzava le attitudini a costruire, a partire da elementi geometrici elementari, forati, secondo un passo costante. Poi è venuto Mazinga e la Barbie. Noi bambini, prima, costruivamo macchine col meccano e poi col lego, poi i bambini impararono a montare e smontare Mazinga, un robot dato, non più costruibile ma mutante..... opportunamente ripiegato c'era anche un Mazinga che diventava orologio. Le bambine impararono a vestire la Barbie, a fargli il bagno, a pettinarla, e, attraverso le Barbie, a subire un modello di donna, bionda americana, con la gamba lunga, modello di bellezza a cui tutte sognavano assomigliare. Il Design Baushaussiano, quello della Braun per intenderci, aveva costruito un modello formale delle merci da imporre al mondo intero, testimonianza della superiorità del moderno occidentale su qualsiasi altra cultura. Era un design geometrico, macchinista, tecnico, astratto ma non minimale, rappresentava la superiore razionalità della macchina, la sua capacità di dar forma ripetitiva e seriale. Anche la Barbie era mondiale ma era di più perché era un modello figurativo, non era una astrazione geometrica ma una morbida carne rosa, seni puntuti, capelli biondi e lunghissimi, era l'immagine seriale di un corpo di donna perfetto e per questo bellissimo, da imporre come il design del moderno al mondo intero [figg. 14-15].

Si poteva far carico delle diversità culturali, nei limiti del vestiario però perché per qualsiasi bambina del mondo la Barbie doveva essere



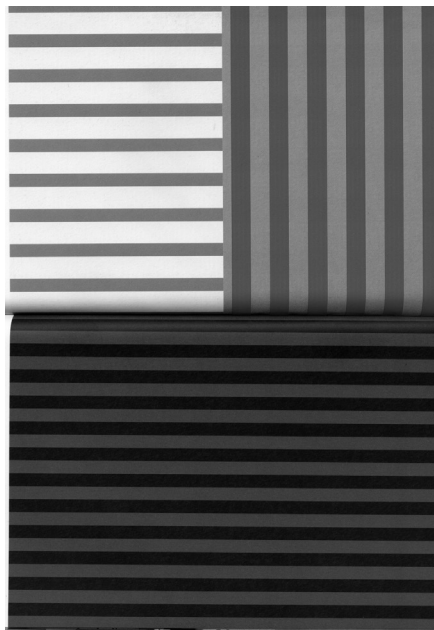
14. Mazinga vs Goldrake

15. Barbie Filippina

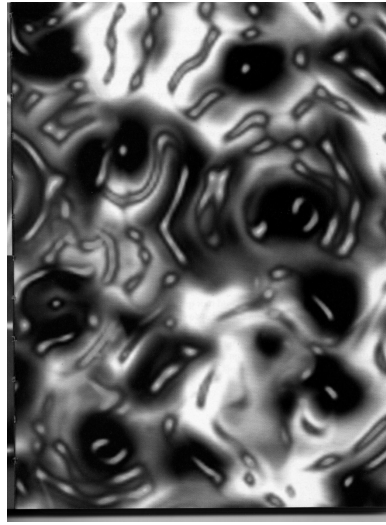
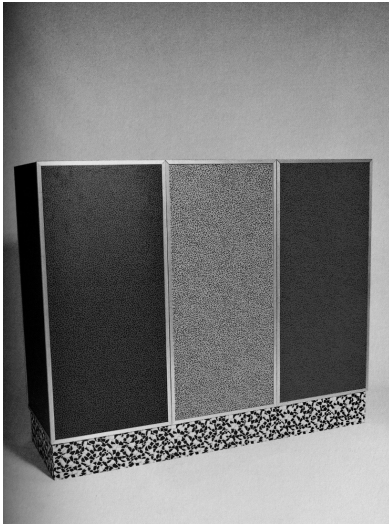
unica, magari vestita con vestiti filippini ma sempre Barbie e americana. Dopo la Barbie, il GameBoy e i giochi elettronici, Lara Croft e un'infinità di modelli da inseguire e far vivere nel virtuale, nell'inconsistenza fisica dell'oggetto, conquistano e strutturano l'immaginario collettivo che si sviluppa nell'infanzia. Cerco di mettere in evidenza come attraverso il gioco siamo sempre stati educati ad identificarci con un progetto di bello, è il primo bello con cui ci formiamo.

Proviamo velocemente un percorso analogo a quello del giocattolo che va dal linguaggio macchinista al virtuale nei laminati di Ettore Sottsass. Quando l'Abet Print negli anni '60 chiese ad Ettore Sottsass di disegnare i suoi laminati Sottsass sostituì alle superfici monocrome e alle texture finto legno la collezione dei rigati. Influenzato forse dai quadri rigati di Frank Stella, Sottsass, ancora nei limiti del geometrico con l'alternanza dei due colori, costruiva una sovrapposizione su due piani della superficie e il laminato non più piatto acquistava una sua profondità [fig. 16].

Negli anni '70 propose dei laminati fatti con i banchi colorati su fondo di altro colore mantenendo l'effetto di profondità dei rigati [fig. 17]. Ma la superficie da geometrica diventò un pulviscolo, si liberò del suo orientamento e rese più indeterminata la percezione del volume e della



16. Sottsass, Laminati rigati, 1965



17. Sottsass, Laminato con banchi, guardaroba, 1979

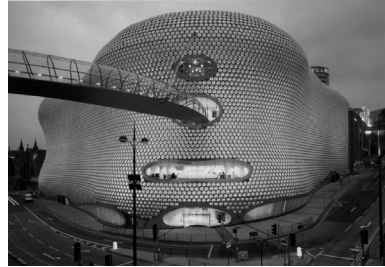
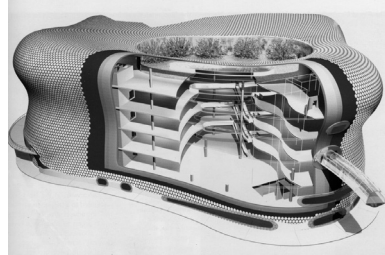
18. Sottsass, Laminato virtual organico, 2000

superficie delle forme geometriche che rivestiva.

Con gli anni 2000 il laminato cominciò a registrare l'immagine del fluido sulla superficie [fig. 18]. La massa materica che viene serigrafata è di una materia viva vista ingigantita al microscopio. Sembra che alcune parti del disegno siano in movimento. Il rigore geometrico di una forma rivestita con questi laminati si dissolve e la figurazione da geometrica diventa fluida organica e virtuale, costruita col computer non più con la fotografia e il disegno. Anche l'architettura ha perso la sua rigidità geometrica, il digitale ha permesso la dissoluzione del geometrico verso l'informe e l'organico e il fantastico. Il Guggenheim Museum di Frank Gery è l'esplosione di un insieme di volumi informi senza geometrie regolari, che si irradiano negli spazi vuoti della città preesistente avvolgendo il viadotto che lo sovrasta, ridisegnando il fiume; un po' polipo, un po' fogli informi portati dal vento, la nuova architettura invade gli spazi della vecchia città di Bilbao riscattandola tutta.

Analogamente il centro commerciale di Future Systems (Jan Kapliky e Amanda Levele) [figg. 19, 20, 21] si va a piazzare nel centro storico di Birmingham, di fronte alla cattedrale gotica, un po' medusa un po' flaccido rospo, ha una pelle formata di specchianti dischi di alluminio satinato che tanto ricordano i vestiti di Paco Rabanne, ma è comunque una pelle, non più certo la raffinata partizione geometrica della superficie di Kuramata, né la facciata lecorbuseriana ordinata dalla sezione

19. Edificio a Birmingham, Future System
 20. Edificio a Birmingham, Future System
 21. Edificio a Birmingham, Future System

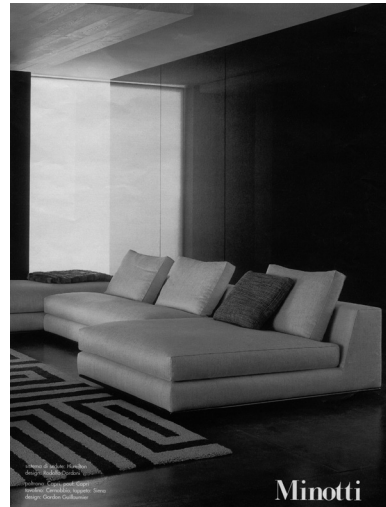


aurea. Questa pelle diventa un involucro plastico inedito rispetto alla tradizionale tettonica degli edifici e nuovo imprevedibile fondale su cui tutte le preesistenze urbane si vanno a ridefinire. Sono queste le «forme impreviste di bellezza che brillano in nuove costellazioni» di cui parlava Bodei? Sono certamente il tentativo di spettacolarizzare l'evento urbano che si chiama architettura. Il fascino di questi edifici è prima di tutto la loro imprevedibilità, rivelano comunque l'acquisita coscienza che quelle verità definitive che avevano affascinato Boullé e che facevano coincidere la bellezza con la perfezione geometrica della forma sono state

superate da una visione evolutiva della scienza che sa di poter contare sempre più su verità non definitive e che sempre più scava non tanto nella scoperta delle leggi universali del cosmo, quanto nelle leggi ancora poco conosciute della vita. È la biologia e non più la fisica la disciplina che guida la corsa alla conoscenza, è la cellula e non più l'atomo l'oggetto della sfida, è il progetto della vita e non più della materia la nuova ambizione dell'umano oggi. Se confrontati con la sfera del cenotafio di Boullè si capisce come nella architettura o meglio in una parte della architettura contemporanea la forma geometrica si sia lentamente mutata in una forma frantumata e organica...e sono le forme di una natura viva, vagamente animale. È cambiato inoltre il rapporto con il pianeta terra e il suo universo, prima una realtà da conoscere, poi da sfruttare e piegare all'infinita voracità degli umani, infine, oggi sempre più vissuta come una realtà da recuperare in un rapporto di compatibilità di fronte al rischio di una distruzione del pianeta che è un'autodistruzione dell'umano. Ma mentre l'architettura sempre meno geometrica sta tentando una radicale ridefinizione del suo essere artificio, l'interior design e il furniture design e una parte della moda, consolidano e diffondono nonostante tutto una netta predilezione per il linguaggio geometrico, astratto e minimale.

Ho inserito due tra le moltissime pagine pubblicitarie di un giornale sull'arredamento di grande diffusione quale «Casa Amica» supplemento mensile del Corriere della Sera per ricordarvi questo dilagare del buon gusto minimale [figg. 22, 23].

Ritorna la composizione delle parti, le forme sono rigorosamente geometriche, sono banditi gli stridori cromatici che tanto affascinavano i radicali, è bandita la figurazione degli oggetti, riassorbita in una stilizzazione che tende all'astratto. Le cassettiere di Kuramata sono un modello di grande qualità formale che più o meno rozzamente viene copiato e reso stile. La B&B azienda italiana che con più coerenza ha cercato di imporre questo stile minimalista è stata acquistata da Bulgari, marchio famoso per i gioielli, oggi impegnato nell'alta moda e nel lusso. L'alta moda con i suoi negozi completamente bianchi e vuoti, riempiti solo da qualche preziosissimo vestito esposto, ha iniziato negli anni Ottanta questo lento riscatto dell'esclusivo rispetto al massificato. La Pop Art con l'intrusione del linguaggio popolare e massificato nel progetto degli arredi, con la valorizzazione del prodotto di massa e l'indifferenza per il prodotto elitario aveva ridotto le differenze di gusto e di linguaggio dei diversi ceti sociali. L'eleganza di Benetton era molto vicina a quella dell'alta moda e c'era un ceto medio in espansione che promuoveva quel gusto. Tutti gli stilisti dell'alta moda si erano messi a disegnare blue



22. Arredi minimal, copertina di «Casa Amica» 2005

23. Arredi minimal, divano, Minotti

jeans, perché il vestiario interclassista aveva preso il sopravvento e la differenza poteva essere solo sottile, ma fin dagli anni Ottanta, e soprattutto negli anni Novanta, si è ricostituito in occidente un forte divario di reddito tra ceti popolari e ceti ricchi, mentre i ceti medi sono entrati in crisi economica e di identità e le differenze tra gusto elitario e gusto di massa si sono ricostruite fino ad essere sempre più marcate. Il gusto popolare contrassegnato da una forte figurazione dei prodotti, finita la fascinazione della massificazione, maturata l'esigenza del ritorno ad una distinzione tra i ceti, è ritornato ad essere sinonimo di cattivo gusto. Il buon gusto, il gusto elitario, si è ammantato di progressiva attenuazione della figurazione fino al recupero dell'astrazione, del geometrico, del minimal, dell'essenziale, dell'archetipico, del bianco e nero e dell'infinità dei toni spenti non chiassosi.

Permangono tuttavia altre idee di bello [fig. 24], il pop è ancora molto presente nel mondo delle discoteche, nella moda giovanile per quanto spesso ibridata dal minimal, ma il minimalismo è sempre più autoritariamente il bello oggi dominante, quello che fa tendenza, che discrimina, che contraddistingue le nuove aristocrazie del gusto. Dopo il relativismo degli anni Sessanta che aveva legittimato il pluralismo estetico è tornata un 'idea' di bello superiore e assoluto. È questo un lungo processo di restaurazione che nelle discipline del progetto iniziò con la XV Triennale di architettura a Milano del 1973. Quella Triennale segnò il trionfo in



24. Stereo Sony Minidisc

Italia dell'architettura 'razionale', quell'architettura che trovava in Boullé il suo antenato e in Aldo Rossi il suo profeta, un architettura rigorosamente geometrica che tentava di imporre le regole della bella forma e che rivendicava una superiorità indiscutibile su qualsiasi altro progetto. Questa architettura accompagna, piaccia o no, l'emergere della cultura neo-conservatrice in tutto l'occidente. Oggi, non a caso, il tema proposto dalle istituzioni cattoliche è la fine del relativismo. Più volte il nuovo Papa si è scagliato contro «il predominio culturale dell'agnosticismo e del relativismo». Gli insegnanti cattolici si stanno battendo per giustapporre il creazionismo al darwinismo nella didattica delle scienze naturali. Un movimento religioso americano «Intelligent Design»³ che si raccoglie intorno al Discovery Institute di Seattle, e che comincia ad avere ascolto anche in Italia, sostiene che «le spiegazioni naturalistiche di alcune entità biologiche non sono possibili e che tali entità biologiche possono essere spiegate solo da cause intelligenti», in altri termini esiste qualcuno che ha progettato l'universo; sono tutti segni convergenti che tendono a riproporre il reale come quella meravigliosa creazione divina che l'umano rappresenta con la rappresentazione del geometrico. Si vuol ritornare alla perfetta sfera di Boullé? Quello che si può dire è che il minimalismo è un linguaggio con precise regole ancorate alle figure geometriche elementari e che nella sua vocazione al rigore ben si coniuga con gli assoluti rivendicati da una chiesa sempre meno tollerante e sempre più reazionaria.

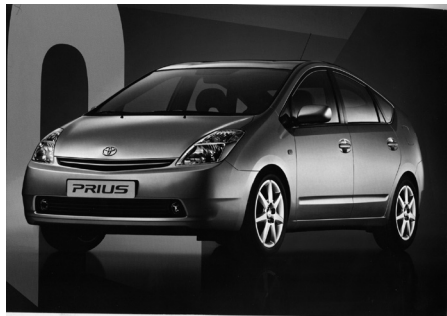
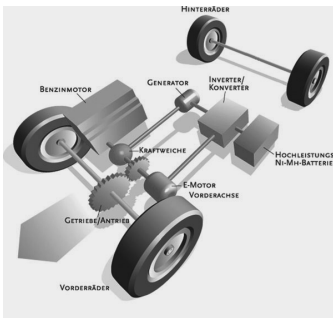
La Toyota ha messo da alcuni anni sul mercato la Prius [fig. 25] una auto di grossa cilindrata che consuma come una citycar perché ha due motori, uno elettrico e l'altro a benzina. Attivando attraverso un computer l'uno o l'altro motore a seconda delle diverse prestazioni richieste all'auto si riesce a realizzare un sensibile risparmio di carburante e di conseguente inquinamento. Il design di questa macchina non è interes-

³ «Micromega», 7/2005.

sante, non è un bel disegno, è bella e innovativa la concezione di questa merce. Non risolveremo più gli infiniti problemi delle diverse società nel mondo con un'unica soluzione; non risolveremo più il problema del vivere né con un'unica tecnologia né con un unico principio estetico. Ognuna delle strumentazioni tecniche a disposizione ha un valore relativo. C'è una strategia culturale che sta alla base dell'ibrido, la coscienza della parzialità dei saperi contro i pensieri unici, la necessità di sinergie di diverse tecniche e diverse culture. A monte di questo progetto della Prius c'è la volontà di ridurre l'inquinamento dovuto alla combustione del petrolio ma anche la convinzione che l'era dei combustibili fossili è finita e che il problema dell'energia troverà soluzione nella progressiva integrazione tra le diverse forme di energia, naturali, solari ed artificiali. La differenza ancora una volta è apparentemente tecnica, in realtà è culturale, e le conseguenze sul piano del bello sono immediate: non più una bellezza costruita sulla perfezione dell'astrazione geometrica coerente con un'ormai superata concezione della bellezza del creato, ma una coniugazione sempre più ricca e articolata di diverse idee di bellezza, espressione di diverse culture, di diversi territori, di diverse ideologie, di diverse religioni, perché non dimentichiamo che è stata sempre la religione l'occasione fondamentale per costruire bellezza. È più bello il tempio greco o la per me stupenda Spianata dei Templi con il Duomo della Rocca a Gerusalemme?

Mi sono regalato alla tenera età di sessant'anni questo giocattolo: Furby [fig. 26], è la bambola della contemporaneità. La Barbie era la perfetta incarnazione della stereotipa bellezza americana, questo giocattolo è un ibrido fatto in Cina: ha gli occhi di un bambino occidentale, il pelo, le orecchie e le zampe di un gatto, sta a cuccia come un cane, parla come un bambino.

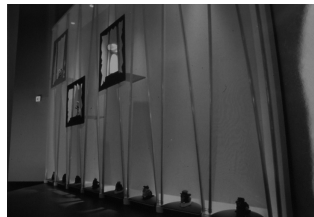
In una mostra a New York (2000) poi ripetuta con poche varianti a



25. Toyota Prius

Tokyo (2001), dal titolo *Esercizi di stile*, organizzata dall'Italian Trade Commission, dalla Federlegno-Arredo e dall'Assopiastrelle, ho costruito in un mio ambiente [figg. 27, 28] questa facciata di plastica e alluminio con alla base, in opportune nicchie, nove Furby, caricati e parlanti. Intorno a me avevo gli ambienti di Studio Cerri e associati, Antonio Citterio e Partners, Paolo Rizzato, Claudio Silvestrin; diverse declinazioni di questa tendenza geometrico minimale che domina l'interior design italiano. I miei Furby densi per la loro figurazione ibrida, giocattolo infantile, in un contesto iper colto, li ho usati, con la solita irriverenza, come faceva Palladio, che usava la scultura per ritmare il disegno della facciata. Il Furby può essere per noi contemporanei quello che era la scultura rinascimentale per Palladio?

Davanti a questa facciata ci sono *Le tre Grazie* omaggio alla scultura dell'armonia ellenistica riproposta dal Canova, sempre irriverentemente da me ridisegnata, non certo in marmo, con tre figure in polistirolo verniciato di acrilico, tre corpi sbraitanti in pose da discoteche vestite da tre progetti fatti da tre mie studentesse dell'ISIA di Firenze: una piccola sfera di ceramica che contiene un po' di terra e un seme che fa crescere una piccola pianta [fig. 29], .è un gioiello povero, 'l'anti Bulgari' da tenere al collo; una tuta attrezzata con resistenze elettriche e computer che produce suono col movimento del corpo [fig. 30]; un calzone terapeutico che attraverso un computer provoca la stimolazione negli arti da



26. Giocattolo Furby

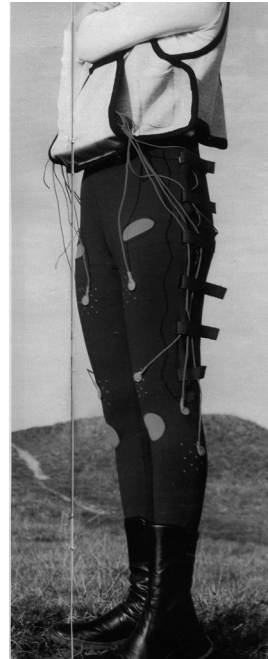
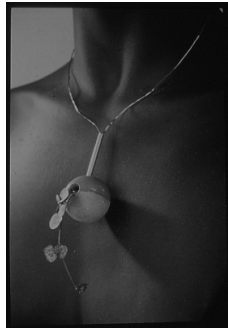
27. Exercise in stile, New York 2000

28. Exercise in stile, New York 2000

riabilitare [fig. 31].

Liberato dall'astrazione geometrica cerco la foma nell'ibridazione del corpo con la merce, utilizzando l'infinita versatilità e la piccola dimensione delle tecnologie elettroniche e tento una sinergia, una compresenza complementare, nella forma naturale con l'artificio costruito partendo dalla ipotesi che una nuova bellezza può essere solo trovata seguendo, accompagnando, le forme naturali invece di squadrarle, geometrizzarle, renderle minimali ed assolute: vorrei fossero vive e relative.

Oggi le cantine per la produzione del vino sono diventate delle spettacolari cattedrali fuori terra, alle volte, in parte, ricoperte di terra riportata. Potentissimi impianti di climatizzazione, con grandi consumi di energia, riescono a ricostruire quel microclima che una volta si trovava nelle cantine interrato. Questa cantina che ho progettato in Spagna insieme al mio committente Carlos Esteve è interamente sottoterra, si adagia sull'andamento dei filoni di roccia sottostanti al terreno che diventano naturali pareti della cantina stessa e l'artificio interamente in cemento armato interessa la copertura dei vani scavati su cui poi verrà riportato il terreno ricostruendo con piccole varianti il disegno del territorio preesistente. Gli impianti di climatizzazione che utilizzeranno energia solare grazie a pannelli fotovoltaici posizionati sulla copertura di terra, sono ridotti al minimo e funzioneranno solo nelle particolari



29. F. Guarneri, *Gioiello*
 30. M. Montinari, *Tregrazie*
 31. P. Perco, *Calzone terapeutico*

condizioni in cui sarà necessario correggere il clima naturale dato dai vani scavati sotto terra lungo i costoni di roccia. Il progetto tende ad essere un ibrido tra naturale ed artificiale, una preesistenza che si dà forma secondo i processi di costruzione della natura che viene rispettata e valorizzata, e un artificiale che integra in un rapporto subordinato il naturale per renderlo funzionale alla produzione e all'abitabilità [figg. 32, 33, 34, 35].

Questa è una lampada in ceramica di nome Geky [fig. 36] progettata per la prossima biennale della ceramica di Albissola che verrà inaugurata nel giugno del 2006.

Ho preso la pancia di un gecko (Tarentule Mauretanica) che si ciba di zanzare e mosche intontite dalla luce elettrica, gli ho messo un filo elettrico su per la coda, che porta corrente ad una lampadina di luce fredda, che consuma poco, posta sopra la sua pancia incisa con una lingua arancione per dare una riflessione calda alla luce fredda, e, per schermare la luce, l'ho protetta con un foglio 'strappato' di policarbonato della Bayer, perché sembri un'ala e perché ci sono gechi che sanno volare.

Geky, mutazione genetica di un filo elettrico in lampada di ceramica



32. P. Deganello, C. Esteva, Pianta cantine Can Rafols dels Caus



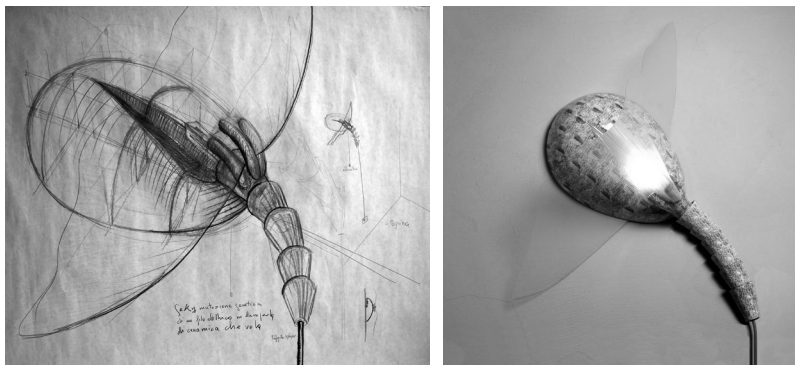
33-35. Foto cantiere cantine Can Rafols dels Caus

che produce luce calda e per di più vola.

Ho realizzato due versioni, ma ne vorrei fare altre, una decorata con la pelle della Tarentula Mauretanica rilevata fotograficamente e rielaborata al computer fino a farla diventare un disegno serigrafabile, l'altra che ripropone la cultura di un fare locale, i decori blu di Albissola.

Il mio progetto oggi lavora sulla valorizzazione e attualizzazione delle culture locali finalizzato ad un recupero di autonomia produttiva dei territori emarginati e/o afflitti dalla fame. Son queste periferie che possono darci quelle forme impreviste di bellezza di cui parla Bodei.

Prediligo forme naturali, animali, cioè vive, di cui cerco una nuova bel-



36. P. Deganello, Disegno lampada Geky

lezza nell'incontro con la tecnica, nella prospettiva di un abitare sempre più rispettoso del naturale, promuovo tecnologie che riducano lo spreco di risorse, ma che non distruggano la figurazione della forma perché cerco sia la più possibile di massa e la più lontana possibile dall'elitaria astrazione e dagli assoluti. Guardo l'arte cercando di capire come disegnare le merci e l'architettura.

Appendice

La guatemalteca Regina José Galindo immerge i suoi piedi in un catino di sangue mestruale [fig. 37] e cammina per le strade del suo paese lasciando le sue orme di sangue, fa un video che fa vedere la ricostruzione del suo imene violentato. Alla Biennale, si è fustigata in pubblico con trecento colpi, quante sono le donne violentate e assassinate nel 2005 nel suo piccolo paese di circa 5 milioni di abitanti. La commissione della Biennale le ha riconosciuto il Leone d'Oro quale migliore artista sotto i 35 anni, «per aver saputo dare vita a una azione coraggiosa contro il potere» per aver fatto sentire sul suo corpo le 300 donne guatemalteche violentate e assassinate in un anno.

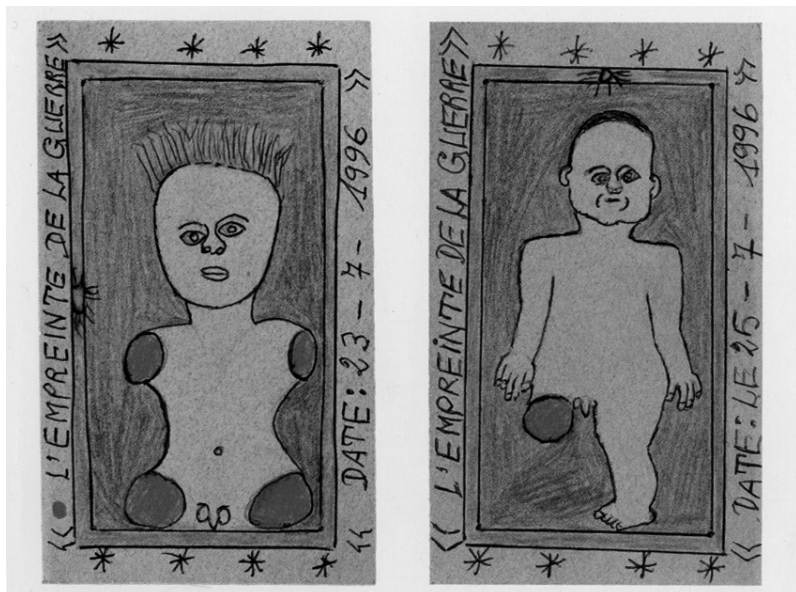
Frederic Bruly Brouabré, artista della Costa d'Avorio fa pitture, grandi come una cartolina [fig. 38], di corpi con gli arti mozzati, costringendo-



37. Performance di Regina José Galindo

ci a vedere gli infiniti massacri che travagliano l'Africa, l'arte fa vedere le mostruosità dell'umano e ci permette sempre meno di volgere lo sguardo dall'altra parte. La realtà luccicante del benessere e dell'abbondanza sembra non attrarre più l'arte che, sempre più, ci fa vedere il conturbante, l'irrapresentabile. A cosa serve quest'arte a noi progettisti se non a farci dubitare che il nostro compito sia sempre e comunque quello di promuovere il consenso sulle merci qualunque sia la ragione del loro essere merci? Rivendichiamo il diritto di dare bellezza solo a merci capaci di essere veicolo di solidarietà tra gli umani e occasione di miglioramento dell'abitabilità di tutto il pianeta, non solo quello dei territori delle nostre vacanze.

Siamo sempre più sottoposti al rischio di un grande nubifragio apocalittico, cerchiamo merci che allontanino questa prospettiva apocalittica.



38. Pitture di Frederic Bruly Brouabré

Finito di stampare nel mese di febbraio 2008
presso le Industrie Grafiche della Pacini Editore S.p.A.
Via A. Gherardesca • 56121 Ospedaletto • Pisa
Telefono 050 313011 • Telefax 050 3130300
Internet: <http://www.pacineditore.it>

