

UNIVERSITÀ DEGLI STUDI DI BOLOGNA
FACOLTÀ DI LETTERE E FILOSOFIA

Corso di laurea in Scienze della Comunicazione

***MODELLI DI RICERCA E DI ESTRAZIONE
DELL'INFORMAZIONE (INFORMATION RETRIEVAL).
ANALISI DI UN'ESPERIENZA SU TESTI MULTIMEDIALI.***

Tesi di laurea in Linguistica Computazionale

Relatore

Presentata da

Prof. ssa Graziella Tonfoni

Carmino Caputo

Cinque parole chiave: Ricerca, Estrazione, Informazione, Multimedialità, Scrittura

Sessione I

Anno accademico 1999-2000

Introduzione

“Cari ragazzi, ricordatevi che aver cultura non significa avere la risposta pronta per ogni interrogativo, ma sapere dove trovare in cinque minuti la risposta al quesito che ci stiamo ponendo”.

Quando, qualche anno fa, ascoltai queste parole per la prima volta dal professor Umberto Eco durante le lezioni del corso di Semiotica, non potevo coglierne appieno la valenza, né tanto meno potevo immaginare quanti studi siano stati fatti e quanto lavoro ci sia ancora da svolgere nel settore degli studi sulla ricerca dell'informazione.

Questo testo si propone di offrire un quadro complessivo generale dell'area di studi che si occupa della ricerca e del reperimento dell'informazione, meglio nota con il termine anglosassone di *information retrieval*, ovvero di mostrarne i progressi fatti sinora e alcune possibili prospettive future. La complessità dell'argomento è dovuta al fatto che si tratta di un'area in cui si intrecciano competenze informatiche, psicologiche, cognitive, semiotiche, linguistiche. La nostra prospettiva sarà fortemente influenzata, in particolar modo, dalla ricerca linguistica, cognitiva e semiotica, ma non potrà trascurare gli aspetti tecnologici del fenomeno. Non potrebbe essere altrimenti, visto che ormai il calcolatore elettronico è divenuto uno strumento fondamentale nel campo della produzione, elaborazione e conservazione del testo. La diffusione ormai capillare degli elaboratori elettronici non solo non esclude la necessità di uno studio di natura linguistica, ma al contrario, richiede a maggior ragione l'intervento di quelle scienze di impostazione umanistica in grado di offrire il loro contributo centrato principalmente sulle necessità dell'uomo piuttosto che sulle possibilità tecnologiche della macchina. Negli Stati Uniti si parla già di una nuova, grande sovradisciplina, chiamata *humanities computing*, volta a comprendere le prospettive dell'utilizzo dello strumento tecnologico e la necessità di rinnovare l'organizzazione del sapere scientifico e non.¹ In Italia un filone più che promettente è quello provvisoriamente definito “informatica umanistica”, con la quale viene indicato “il complesso di questioni teoriche e metodologiche poste a chi si occupa di discipline ‘umanistiche’ con una strumentazione informatica”².

¹Paola Spina, *Parole in rete*, Firenze, La Nuova Italia 1997.

²Alberto Cadioli, *Il critico navigante [saggio sull'ipertesto e la critica letteraria]*, Genova, Marietti, 1998, p.10.

L'idea alla base del nostro lavoro è che le modalità di produzione della scrittura nell'era della comunicazione globale, indipendentemente dal supporto fisico utilizzato per la diffusione, debbano prendere nella dovuta considerazione il problema del reperimento dell'informazione da parte del destinatario. Non si può celare il fatto che soprattutto negli ultimi decenni si sia avvertita un'urgenza nel perfezionamento e nell'ampliamento di sistemi di catalogazione, indicizzazione, categorizzazione, dovuta all'enorme quantità di informazioni di vario genere di cui il progresso tecnologico ha reso possibile la produzione. Il 'fenomeno Internet' ha messo sotto gli occhi di tutti, utenti e ricercatori, il drammatico bisogno di strumenti efficaci di sintesi, elaborazione e ricerca dell'informazione: milioni di documenti in rete sono a disposizione di chiunque sia in grado di connettersi, a patto di trovarli. Al di là delle barriere tecnologiche, economiche e di alfabetizzazione informatica che, per quanto di decisiva importanza, in questa sede non verranno analizzate, altre e notevoli barriere ostacolano il raggiungimento dell'informazione desiderata: la scarsa visibilità e rinvenibilità dei documenti. A tutti i navigatori è capitata, prima o poi, l'esperienza di sentirsi impotenti di fronte a migliaia di documenti indicizzati, tra i quali tre o quattro sono quelli che ci interesserebbero da vicino, e le altre migliaia sono assolutamente inutili. Come ricorda Paola Spina nella sua guida ai siti Internet sul linguaggio,

L'attività di *information brokerage*, la capacità, cioè, di selezionare il valore dell'informazione in relazione alle peculiari necessità di un gruppo di potenziali fruitori e di raccoglierle in documenti facilmente consultabili, magari sotto forma di guide, cataloghi o indici, costituiscono, nell'era della globalizzazione della conoscenza, uno degli investimenti più produttivi.³

Un altro dei temi che trasversalmente ritornerà in questo testo è quello, se vogliamo di senso contrario, per cui è vero che la diffusione di Internet ha reso più evidente il bisogno di studiare dei sistemi euristici avanzati; ma è anche vero che questo bisogno, magari in maniera meno evidente, non è certo nuovo; né sono necessariamente tutti nuovi gli strumenti necessari al suo soddisfacimento. Ci si oppone qui, insomma, a quella tesi, contestata anche da Volli che la accusa di determinismo tecnologico, per cui il succedersi

³Paola Spina, op.cit.,p.X.

delle tecniche porta a cambiamenti epocali rivoluzionari. La tecnologia ‘dura’, quella delle macchine, è solo uno degli elementi da considerare per analizzare il cambiamento, sicuramente è un elemento molto rilevante, ma deve fare i conti con una serie di tecnologie ‘molli’ costituite da grammatiche, sistemi di regole e convenzioni, per non parlare di variabili sociali ed economiche. La determinazione, conclude Volli, che è comunque graduale e comporta una lenta evoluzione più che una radicale trasformazione, è storica, non tecnologica, perché “non sono le nuove tecnologie a prevalere necessariamente, ma le tecnologie adatte rispetto a certi fondamentali fini sociali e ai contenuti che essi comportano”⁴. Detto ciò, si comprende lo sguardo attento al passato e alla storia che deve caratterizzare un lavoro che si occupi di sapere umanistico e tecnologia. I principi dell’euristica, dell’indagine bibliografica, del reperimento di informazioni necessarie, infatti, non nascono certo oggi, dal momento che l’utilizzo di supporti fisici volti alla conservazione delle informazioni (per quanto possa apparire sgradevole questa definizione, essa si adatta benissimo all’idea di libro) ha da sempre comportato l’uso dei sistemi di catalogazione, indicizzazione, recupero. Si pensi, per esempio, ad una forma di reperimento dell’informazione che non può certo dirsi recente, come l’utilizzo dei versetti per l’indicizzazione dei brani dell’Antico Testamento. Essa risale alle scuole giudaiche dei primi secoli dopo Cristo, anche se fu san Girolamo a fissare il canone della ‘Vulgata’, traducendolo direttamente dall’ebraico. A lui si deve una prima suddivisione in ‘capi’ sia sul testo greco sia su quello latino. La divisione in versetti del Nuovo Testamento risale a Sante Pagnini ed è del 1527.⁵ È a nostro parere indispensabile, lo ripetiamo, essere capaci di guardare alla storia, o alla preistoria, dell’*information retrieval*, per comprenderne le basi cognitive e logiche, e, solo in un secondo momento, cercare di mettere in atto le implementazioni tecnologiche. La breve digressione operata sulla divisione in versetti del Libro per eccellenza ci ha fornito infatti materia per due spunti interessanti: il primo è che l’utilizzo di rapidi strumenti di consultazione del testo diventa tanto più ingente quanto più la natura del testo si presta ad una lettura non lineare, come nel caso del testo biblico; il secondo è che, per quanto possano ritenersi utili ed efficaci, tali strumenti non sono mai assoluti, ma sono il risultato di una decisione, di una presa di posizione arbitraria e quindi contestabile: le traduzioni bibliche,

⁴ Ugo Volli, “Nuove e vecchie tecnologie”, in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.164.

⁵ Enciclopedia Treccani, vol. VI.

oggi, non riportano tutte la stessa suddivisione in versetti.

Nei discorsi riguardanti le nuove tecnologie, l'informatica, l'intelligenza artificiale, scrive Gigliozzi, uno dei problemi che si pone più di frequente è quello di mettersi d'accordo sulla terminologia da utilizzare. Gigliozzi propone allora un punto di partenza comune, una base sulla quale è possibile poi costruire un discorso sulle nuove tecnologie, che noi condividiamo pienamente:

Qui possiamo scegliere una scorciatoia e stabilire l'equivalenza nuove tecnologie = informatica. A patto però che prima ci si accordi in modo da intendere, etimologicamente, l'informatica come trattamento automatico delle informazioni e non come disciplina che ci aiuta a far funzionare il computer. Stabilito questo tutti gli altri tasselli vanno al loro posto. L'information retrieval, le basi di dati (testuali e non), la multimedialità, la telematica, persino la struttura stessa dell'elaboratore e via dicendo, tutto questo può essere letto nei termini di trattamento (acquisizione, conservazione, trasmissione, elaborazione) dell'informazione.⁶

Gigliozzi accenna ad una possibile definizione di *information retrieval*, considerandolo uno dei possibili trattamenti dell'informazione; è opportuno allora cercare di approfondire la definizione di questo campo di studi, dei suoi fini, dei suoi mezzi, delineandone le differenze con un'altra area molto importante, l'*information seeking*. Già la traduzione italiana, 'reperimento' o 'ritrovamento' dell'informazione piuttosto che 'ricerca', ci permette di cogliere alcune diverse sfumature di significato. Premesso che qui per informazione intendiamo, con la definizione molto ampia di Marchionini⁷, tutto ciò che può cambiare il bagaglio di conoscenze di una persona, bisogna considerare che quando parliamo di reperimento, ci riferiamo ad un oggetto di cui è già nota l'esistenza, che qualcuno ha già catalogato e posizionato per renderlo fruibile a qualcun altro; l'*information seeking*, invece, connota il processo di acquisizione della conoscenza, che può essere efficace e

⁶ **Giuseppe Gigliozzi**, "Critica letteraria e nuove tecnologie", in Patrizia Bellmann Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.56.

⁷ **Gary Marchionini**, *Information Seeking in Electronic Environments*, Cambridge, University Press, 1995, p.5.

rinvenire un oggetto oppure no. Sempre secondo Marchionini, l'*information retrieval* trova applicazione dove c'è un *database*, una struttura più o meno organizzata di dati; l'*information seeking*, invece, è più strettamente connesso al processo dell'apprendimento⁸. Questa distinzione non è superflua ai nostri fini, perché oggetto di interesse per il nostro studio è proprio l'operazione attraverso la quale un produttore di testi organizza l'informazione e la dispone in maniera tale da essere facilmente fruibile. Dunque, in estrema sintesi, possiamo dire l'*information retrieval* si occupa delle tecniche di reperimento dell'informazione da una base di dati che è già stata in qualche modo elaborata.

Dopo aver presentato le basi più decisamente linguistico-cognitive dell'*information retrieval*, questo lavoro si occuperà di definire il campo della ricerca testuale odierna, che come sappiamo, non si sviluppa più soltanto sui testi cartacei, ma anche su quelli elettronici. Se è vero che il computer ha cambiato la nostra modalità di scrittura, è altrettanto vero che questo cambiamento deve favorire il successivo ritrovamento dell'informazione, e non complicarlo. Il testo elettronico, infatti, con la possibilità di integrare linguaggi di scrittura tradizionalmente separati, mostra una ricchezza e una complessità sconosciuta sinora, che può facilmente generare fenomeni di disorientamento se non abilmente gestito.

Ci occuperemo inoltre degli strumenti informatici a disposizione di chi, oggi, si ponga alla ricerca di determinate informazioni, cercando di confrontare i progressi tecnologici fatti con i risultati ottenuti dalle scienze linguistiche.

La conclusione, che possiamo già per sommi capi anticipare, di tale disamina ci porterà ad affermare il bisogno da parte dei produttori di informazioni di prendere atto della complessità del quadro nel quale il loro testo si andrà ad inserire, e di conseguenza si avvalorerà la necessità, anche con l'ausilio degli strumenti informatici, di strutturare l'informazione in maniera tale da favorirne la fruizione. Le capacità di sintetizzare, comprimere, riassumere un testo diventano fondamentali per il produttore di informazione dell'era della rete globale. E poiché il presente elaborato è soprattutto quello che in ambiente anglosassone si definirebbe un *case-study*, l'analisi di un'esperienza, tratteremo appunto di un'esperienza in cui si è progettata una confluenza di strumenti mediatici e di 'filtro editoriale' con il

⁸ Gary Marchionini, op.cit.,p.6.

fine di semplificare il processo di apprendimento e ricerca da parte dell'utente. Sulla scorta di quanto già in precedenza evidenziato, si descriverà un'iniziativa editoriale caratterizzata dal tentativo di mettere in atto alcune delle teorie dell'*information retrieval* e, più in generale, dell'*information design*, inteso come progettazione del testo e pianificazione del processo di scrittura. Si cercherà di definire l'insieme di operazioni attraverso le quali si è operata una selezione di brani da quattro testi per arrivare ad un quinto testo 'estratto' dai precedenti, considerando le funzionalità di un supporto multimediale, e giudicandone i risultati rispetto alle aspettative.

Sarà possibile infine trarre un bilancio del lavoro realizzato e valutare l'efficacia di tale progetto attraverso la verifica delle opinioni e delle esperienze di un gruppo di studenti che hanno utilizzato il volumetto come supporto di apprendimento nell'ambito di un progetto di *distance learning* attivato fra l'Università di Bologna e la Facoltà di Architettura del Politecnico di Milano.

1.L'Information Retrieval: dalla prospettiva cognitiva a quella computazionale

"Bisogna che tu lo sappia: in questo basso mondo tutto comincia con l'immagine e prosegue con la metafora. Il significato sta a te conquistarlo, con la forza del neurone!" (D. Pennac)

1.1. Il punto di vista della linguistica computazionale

L'approccio privilegiato che abbiamo seguito nel corso della ricerca è stato quello linguistico computazionale, per cui non pare inopportuna qui una breve premessa sugli ambiti della disciplina, molto più ampi di quanto non possa apparire in prima istanza e ancora non rigidamente definiti, vista la relativa giovane età della stessa.

La linguistica computazionale è, infatti, una scienza piuttosto 'giovane' e che potremmo definire 'di frontiera'. Di frontiera perché sperimenta tecniche e modelli che non possono contare su una tradizione consolidata, ma solo sull'intuito e l'entusiasmo di chi sperimenta sapendo di essere un pioniere. Di frontiera anche perché essa si pone ai margini dell'area linguistica, e umanistica in genere, rivolgendosi verso settori tradizionalmente di competenza ingegneristica e informatica. L'uso dei termini 'aree', 'margini', 'frontiere', potrebbe apparire inappropriato o comunque obsoleto, dal momento che l'interdisciplinarietà e la confluenza di saperi sono da tempo parole d'ordine nelle scienze della comunicazione. Tuttavia una tale suddivisione un po' rigida e, se vogliamo, arbitraria, ci permette di comprendere il retroterra culturale da cui una disciplina parte e di conseguenza gli aspetti che prenderà maggiormente in considerazione. Dunque, la linguistica computazionale parte da un'area umanistica e si rivolge ad una informatica: questa connotazione 'vettoriale' si mostrerà utile ai fini della comprensione della natura della materia. Essa si occupa

principalmente di analizzare i modelli di elaborazione testuale sviluppati attraverso l'uso del computer: un'area molto complessa che non si limita al linguaggio ma all'informazione nel suo complesso. In effetti quelle linguistiche sono solo alcune delle numerose competenze richieste dal processo di elaborazione dell'informazione. La linguistica computazionale, non vada dimenticato, è innanzi tutto 'linguistica', e come tale risente delle due tradizioni che caratterizzano questo tipo di studio: da una parte, linguistica come studio del linguaggio, inteso come facoltà cognitiva e comunicativa dell'individuo; dall'altra, linguistica come studio della varietà linguistica, delle lingue intese come manifestazioni diverse. La linguistica computazionale si muove in entrambe le tradizioni: è lo studio, l'area di discussione e di analisi di modelli rigorosi, ma considera anche la varietà del mondo linguistico, fermo restando il predominio degli studi sulla lingua inglese. L'aggettivo 'computazionale' indica la tendenza della materia ad evidenziare l'aspetto numerico, la quantificazione degli aspetti salienti del linguaggio naturale sulla base di rilevazioni di frequenza di natura statistica. È questa l'accezione tradizionale della linguistica computazionale, quella più nota. Si tratta tuttavia di un'accezione ristretta che oggi andrebbe allargata, tanto che si arriva a parlare frequentemente di 'linguistica informatica', di studio dei modelli computazionali e informatici da adattare allo studio del linguaggio naturale.

Gli studiosi di linguistica computazionale adattano sovente modelli linguistici all'elaborazione elettronica, mettendoli alla prova con problemi computazionali; ma operano anche un procedimento inverso: arrivano all'analisi del linguaggio partendo da modelli tipici della cibernetica. In ogni caso il risultato deve essere operativo, verificabile, cioè 'deve funzionare'. Dietro la trasformazione 'terminologica' della linguistica computazionale in linguistica informatica si cela il tentativo di offrire un quadro complessivo più esaustivo, inserendo l'elemento qualitativo accanto al tradizionale aspetto quantitativo. La linguistica diventa qualitativa perché non si occupa più semplicemente di quantità di elementi, ma di significazione, con la conseguente riorganizzazione delle priorità e l'avvicinamento alle scienze come la semiotica o la filosofia del linguaggio, che hanno da sempre lasciato molto spazio agli aspetti semantici dello studio del linguaggio.

La connotazione 'vettoriale' della linguistica computazionale, il suo partire da basi tradizionalmente umanistiche per muoversi nella direzione di studi più marcatamente informatici, la distingue da una branca di studi affine che si

differenzia per una connotazione vettoriale opposta: è l'informatica linguistica, che processa il linguaggio per fini esclusivamente informatici, cercando nel patrimonio linguistico elementi che possano essere funzionali agli scopi dei programmatori, mutuando teorie e modelli propri della linguistica che si adattano alla risoluzione di problemi informatici, quali la costruzione di interfacce in linguaggio naturale, sofisticate ma 'user-friendly', facili da usare.

1.2. Alcuni principi semiotici rilevanti

Alcuni principi fondamentali della teoria semiotica saranno continuamente presi come riferimento nel corso di questo lavoro. Essi, pur rimanendo talvolta sullo sfondo del discorso elaborato, andrebbero comunque sempre tenuti presenti durante la lettura, anche perché non sarebbe possibile menzionarli ogni qual volta se ne presentasse l'occasione. Pare opportuno pertanto soffermarci qui, una volta per tutte e in maniera sintetica, benché la complessità dell'argomento richiederebbe maggiori approfondimenti, su alcuni elementi teorici elaborati in particolare da Luis Hjelmslev, che a sua volta riprendeva e perfezionava alcune idee che già erano state proposte da Saussure.

Sostiene giustamente De Mauro⁹ che la maggior parte delle tendenze teoriche in ambito linguistico sviluppate nella seconda metà del ventesimo secolo devono molto, in termini di influenza e di ispirazione talvolta decisive, alla meditazione dello studioso danese. Sarebbe inutile fare un elenco degli studi successivi che hanno tratto spunto dalle osservazioni di Hjelmslev: praticamente tutte le correnti di pensiero linguistiche hanno dovuto fare i conti prima o poi con le sue affermazioni, anche se in maniera non sempre dichiarata, e, chissà, qualche volta persino inconsapevole. Come dicevamo in precedenza, il discorso di Hjelmslev parte dall'analisi del *Cours* di Saussure, che anzi probabilmente, sostiene ancora De Mauro, non avrebbe goduto della fortuna che gli è stata attribuita se non fosse stato per la rilettura fruttuosa che ne fece Hjelmslev nella sua opera più famosa, i *Fondamenti della teoria del linguaggio*, nel 1943. Per quanto concerne il nostro lavoro sugli strumenti per il reperimento dell'informazione, siamo senz'altro debitori nei confronti di Hjelmslev quando riproponiamo uno degli aspetti metodologici maggiormente difesi dallo studioso danese: l'attribuzione di rilevanza all'aspetto semantico dei segni e dei linguaggi. Tale aspetto è

⁹ Si veda l'introduzione a **Luis Hjelmslev**, in *Saggi Linguistici*, Bari-Roma, Laterza, 1981.

purtroppo assai spesso trascurato, come avremo occasione di approfondire in seguito, nei sistemi di ricerca automatizzati. Gli elementi della teoria hjelmsleviana che recupereremo nel corso di questa analisi sono soprattutto quelli che si rivolgono alla definizione di lingua: per Hjelmslev una lingua è una forma specifica organizzata fra due sostanze, quella del contenuto e quella dell'espressione. Nonostante la lingua ci appaia fundamentalmente come un sistema di parole e di segni, cioè come un sistema di espressione unito a un relativo contenuto, sarebbe più corretto considerarla un sistema di elementi dotato di regole per la loro combinazione, sia nel contenuto che nell'espressione. Si tratta di un sistema formale, indipendente dalla sostanza: per riportare un esempio dello stesso Hjelmslev, l'italiano parlato, l'italiano scritto, l'italiano telegrafato mediante il codice morse, l'italiano trasmesso con le bandierine secondo il codice internazionale della marina, resta pur sempre la stessa lingua, nonostante le manifestazioni concrete siano differenti. La lingua è una forma, un sistema di relazioni tra unità, indipendente dalla sostanza, cioè dal concreto realizzarsi fisicamente delle unità. L'idea tradizionale di segno, in Hjelmslev, è messa in crisi, e sostituita da quella di funzione segnica: quest'ultima è una relazione esistente tra elementi di contenuto ed elementi dell'espressione definiti funzionalmente, tra i quali, però, non esiste corrispondenza biunivoca, anche se i segni sono scomponibili in elementi più piccoli. Inoltre, ed anche in ciò si nota l'eredità saussuriana, il rapporto tra forma e sostanza è arbitrario, così come quello tra espressione e contenuto. Ricapitolando: la lingua è un sistema di relazioni formali, indipendenti dalle sostanze, e si compone di due piani, il piano dell'espressione e il piano del contenuto. Questi due piani sono coesistenti e non alternativi, sono interdipendenti, e sono posti in relazione dalla funzione semiotica, o segnica. Di ogni enunciato, pertanto, possiamo identificare un contenuto, costituito da una forma linguistica e da una sostanza, e un'espressione, anch'essa costituita di una forma che 'ritaglia' la sostanza. Hjelmslev definisce il piano dell'espressione un piano esterno, costituito da sostanza fisica; laddove il piano del contenuto è un piano interno, sostanza ontologica. Quando un cambiamento tra unità di espressione provoca un cambiamento tra unità del contenuto, o viceversa, si ha commutazione: attraverso questa prova non solo si colgono le differenze formali tra i due piani, ma è possibile identificare gli elementi minimali che li costituiscono¹⁰.

¹⁰ La ricerca delle unità minimali di contenuto, in realtà, si è finora dimostrata vana; ma non è il caso di approfondire tutte le definizioni di Hjelmslev e i loro limiti, il che supera gli scopi che si prefigge questo paragrafo, e cioè una sintetica inquadratura di una teoria che meriterebbe ben altri spazi.

Se la commutazione, che si genera tra invarianti, non si realizza, si parlerà di semplice sostituzione, possibile tra varianti.

Le unità del piano dell'espressione sono, per la loro stessa natura 'esterna' e 'fisica', individuabili abbastanza facilmente, così com'è possibile individuarne le loro relazioni attraverso l'elaborazione di una sintassi. Ben più problematica è invece la gestione del piano del contenuto, delle sue relazioni ontologiche, degli aspetti semantici. Per quel che riguarda la ricerca dell'informazione, possiamo anticipare già da ora che 'le macchine', dal più artigianale strumento meccanico al più sofisticato sistema esperto, sono in grado di muoversi solo sul piano dell'espressione: riconoscono occorrenze fisiche, ma non sono in grado di interpretarle, di fornire loro un significato, di elaborare cioè le corrispondenti unità del contenuto.

1.3. Un contributo semantico alla risoluzione di alcuni problemi computazionali

Ci pare di dover sgombrare il campo, a questo punto, dal rischio di considerare, equivocamente, l'approccio linguistico computazionale come un approccio rivolto esclusivamente alla catalogazione statistica di occorrenze, concentrato, dunque, sul puro livello dell'espressione. Al contrario, l'area linguistico computazionale in cui la nostra ricerca si muove è fortemente influenzata dai risultati della linguistica testuale, che in tempi più recenti rispetto a Hjelmslev, e con occhio rivolto agli sviluppi e alle possibilità della tecnologia, ha riscoperto e valorizzato il contributo semantico per gli studi del linguaggio. Per i linguisti testuali, infatti, il testo è, con le parole di van Dijk, "l'unità complessa minima sintatticamente, semanticamente e pragmaticamente completa", per cui l'aspetto semantico e quello pragmatico (rifacendoci qui alla tradizionale e per alcuni anacronistica tripartizione di Morris¹¹) non possono essere in alcun modo trascurati. Il testo differisce dalla frase per il grado di complessità e non per le dimensioni quantitative, essendo, un'unità minima, scientificamente distinguibile e complessa, cioè tridimensionale. Il testo è un'occorrenza comunicativa che deve rispettare sette requisiti per essere considerato tale: la coesione, la coerenza, l'intenzionalità, l'accettabilità, l'informatività, la situazionalità (*situationality* in inglese) e l'intertestualità¹². In altri termini gli elementi che

¹¹ **Charles Morris**, *Foundations of the Theory of Signs*, Chicago, University of Chicago Press, 1938.

¹² **Robert Alain de Beaugrande, Wolfgang Ulrich Dressler**, *Introduction to Text Linguistic*, London, Longman 1981.

compongono un testo, perché questo realizzi la sua potenzialità comunicativa, devono essere mutualmente connessi in una sequenza, devono rappresentare concetti e relazioni ‘sottostanti’ al testo accessibili e rilevanti, e tutto ciò deve rispondere ad un preciso ed intenzionale progetto del produttore volto al raggiungimento di un obiettivo. Tale progetto (*plan*) richiede la collaborazione del ricevente, che deve accettare che gli elementi elaborati dal produttore siano coesi, coerenti, e per lui rilevanti. Il testo deve essere inoltre informativo, cioè presentare occorrenze inaspettate, aumentare la conoscenza del ricevente, ed essere legato inequivocabilmente al contesto situazionale in cui è espresso, oltre che dipendente dalla conoscenza generata eventualmente da altri testi. La rilevanza del contesto comunicativo in cui il testo trova espressione è uno dei concetti chiave di questa ‘scuola’¹³: “units of content are not fixed particles with a stable identity, but rather fuzzy agglomerates sensitive to the conditions of their usage”¹⁴. Come ricorda Mary Elizabeth Conte¹⁵, numerosi fenomeni linguistici non possono spiegarsi in riferimento a un enunciato singolo, senza prenderne in considerazione il contesto situazionale, così come la coerenza stessa del testo non può considerarsi indipendente dal contesto pragmatico in cui il testo viene prodotto e percepito. Fattori quali parlante, ascoltatore, luogo e tempo del discorso influenzano la coerenza semantica globale di un testo. Questo approccio volto a considerare le condizioni ambientali in cui un testo si manifesta ha delle valenze importantissime quando si prende in considerazione l’*information retrieval*, perché in nessun caso si può pensare ad un efficace sistema di ricerca dell’informazione senza prima aver considerato la modalità secondo la quale esso prenderà forma, l’utente a cui si rivolgerà, i tempi e i luoghi di fruizione. E i criteri meramente quantitativi e statistici, che non considerino anche i contenuti testuali, mostrano, come vedremo, tutti i loro limiti quando si ha a che fare con grosse porzioni di testo. Inoltre l’approccio della linguistica testuale si definisce ‘procedurale’: gli autori più rappresentativi, denunciando la trascuratezza con cui alcuni linguisti hanno relegato lo studio dell’uso del linguaggio nel dominio marginale della pragmatica, propongono una maggiore attenzione a questo

¹³ Il termine ‘scuola’, con tutto ciò che da esso si può presagire, mal si adatta alla definizione di questo gruppo di studiosi: van Dijk stesso sottolinea il fatto che le il termine ‘linguistica testuale’ non può designare una teoria o un metodo singolo, quanto piuttosto ogni lavoro di linguistica che prenda in considerazione il testo come primario oggetto di analisi. Si confronti a riguardo **Robert Alain de Beaugrande, Wolfgang Ulrich Dressler**, op.cit., p.15.

¹⁴ **Robert Alain de Beaugrande, Wolfgang Ulrich Dressler**, op. cit., p. 22.

¹⁵ **Mary Elizabeth Conte** (a cura di), *La linguistica testuale*. Milano, Feltrinelli, 1977, pp.14-16.

settore. La pragmatica e i problemi relativi all'uso del linguaggio non sono indifferenti all'area sintattica e semantica, soprattutto nella misura in cui riscoprono il ruolo, fondamentale per i linguisti testuali, del progetto (*plan*) e dell'obiettivo (*goal*) comunicativo. Un approccio 'procedurale' si concentra sulle operazioni che manipolano le unità e i modelli durante l'utilizzo dei sistemi linguistici, più che sulla scoperta di queste unità e di questi modelli, che pur restando attività d'analisi e investigazione, non sono obiettivi in sé. Lo studio dei sistemi di reperimento dell'informazione che verrà in seguito proposto sarà procedurale, nella misura in cui si sforzerà di prendere in considerazione le procedure attivate dall'utente durante il suo percorso attraverso il testo alla ricerca dell'informazione rilevante, seguendo un progetto, più o meno conscio, volto alla realizzazione di un obiettivo.

Tornando all'enunciato di sopra in cui si sottolineava l'importanza dell'operatività dei risultati della linguistica computazionale, tale operatività non si traduce semplicemente nella costruzione di modelli verificabili empiricamente, ma, piuttosto, nella costruzione di una teoria della conoscenza che permetta una efficace descrizione delle modalità in cui questa conoscenza è applicata durante le attività comunicative. Anche in questo caso il 'debito' nei confronti della linguistica testuale è evidente: in un saggio, Petöfi, dopo aver spiegato la non operazionalità della tripartizione tradizionale della semiotica verbale (sintassi, semantica e pragmatica), e dopo aver messo in luce come una teoria complessiva linguistica non possa escludere dal proprio dominio gli aspetti logici, sociologici, comunicativi, psicologici non direttamente analizzabili da una grammatica concentrata sulla sintassi, evidenzia che, ai fini di una elaborazione di una teoria testuale,

Il primo e principale compito dovrebbe essere descrivere la conoscenza concernente le strutture testuali (in modo tale, ovviamente, che questa descrizione consenta la descrizione dell'applicazione comunicativa di questa conoscenza, cioè della produzione e recezione dei testi, in modo compatibile con la descrizione della conoscenza stessa).¹⁶

La descrizione di una struttura testuale, dunque, può dirsi valida solamente se è in grado di descrivere efficacemente la produzione e la ricezione dei

¹⁶János S. Petöfi, "Semantica, pragmatica, teoria del testo", in Mary Elisabeth Conte (a cura di), *La linguistica testuale*, Milano, Feltrinelli, 1977, p.214.

testi: per un linguista computazionale qualsiasi teoria va testata, sperimentata, provata, e se, per quanto elegante, internamente coerente e apparentemente solida, non si dimostra efficace, va accantonata, per quanto doloroso ciò possa risultare.

1.4. La logica del ragionamento comune

Quanto detto sinora ci permette di introdurre un altro aspetto che caratterizza gli studi di linguistica computazionale, e, conseguentemente, il presente elaborato: la consapevolezza dell'insufficienza degli strumenti della logica classica nei confronti di un oggetto di analisi, il testo, che negli anni si è dimostrato più complesso di quanto non si ritenesse in un primo momento. È stato infatti ormai ampiamente dimostrato come un approccio logico che presupponga un apparato di regole sistematico può rivelarsi completamente avulso da qualsiasi tipo di contesto comunicativo reale, e, quindi, fundamentalmente inutilizzabile. Ballmer¹⁷ spiega come l'ingresso di numerose discipline come l'intelligenza artificiale o la psicologia in campi, quali quelli dell'analisi del linguaggio, tradizionalmente gestiti dai logici, abbia procurato non poche difficoltà al paradigma vigente. Se infatti era un tempo generalmente condivisa l'opinione secondo la quale le espressioni linguistiche descrivevano la realtà, per cui potevano essere suddivise in vere o false rispetto a questa realtà, l'idea che il linguaggio potesse oltrepassare questa realtà, o modificarla, ha complicato il quadro. Si pensi alle differenze tra il cosiddetto primo e secondo Wittgenstein¹⁸, un autore la cui ricerca è un segno emblematico di questa svolta, o alla 'rivoluzione' apportata dalla teoria degli atti linguistici. Ballmer continua evidenziando come

Le espressioni linguistiche possono allora essere prese al massimo come espressioni di modelli, il che non significa necessariamente che siano modelli essi stessi di qualcosa di reale.¹⁹

e confutando l'idea che sia possibile la diretta traduzione delle espressioni linguistiche in espressioni logiche rappresentanti condizioni di verità. Non è qui il caso di approfondire un argomento che di per sé richiederebbe numerose ricerche e approfondimenti. Basti qui poter dire che un approccio

¹⁷ **ThomasT. Ballmer**, *Frames e strutture contestuali*, Parma, Ed.Zara, 1984.

¹⁸ Cfr. in particolare **Ludwig Wittgenstein**, *Tractatus logico-philosophicus*, Torino, Einaudi, 1964, e *Ricerche filosofiche*, Torino, Einaudi, 1967.

¹⁹ **ThomasT. Ballmer**, op.cit., p. 12-13.

linguistico computazionale, o, più generalmente, un approccio moderno all'*information retrieval* non può limitarsi a prendere in considerazione le regole della logica classica. La ricerca cognitiva, in questo campo, dimostra di essere molto più avanti, come vedremo, rispetto a quella tecnologica: i cosiddetti motori di ricerca del *web*, infatti, interagiscono con l'utente attraverso un linguaggio piuttosto rozzo basato sui connettivi della logica booleana, quando sarebbe opportuno sviluppare una modellistica improntata su modalità di ragionamento più 'umane'.

Un'area di sviluppo della logica che invece pare molto promettente anche se è ancora presto per dire con esattezza quale contributo potrà apportare ai processi di *information retrieval* è quella del *common sense reasoning*, la logica del senso comune. È bene chiarire subito che il ragionamento del buon senso è di gran lunga più complesso del ragionamento logico formale, come sostenuto da Minsky: "La logica richiede un solo sostegno per quanto lungo sia il legame, un'unica deduzione impeccabile. Il buon senso chiede a ogni passo se tutto ciò che si è trovato fino a quel punto è in accordo con l'esperienza quotidiana"²⁰. Il modo di ragionare basato sul buon senso differisce dal modo di ragionare logico, perché nella vita quotidiana non ci accontentiamo di catene di ragionamenti efficaci, ma cerchiamo continuamente di cercare altre prove oltre a quelle fornite dall'argomentazione logica. Quest'ultima può servire semmai non tanto a fornire idee nuove, quanto a scoprire la debolezza di quelle vecchie, essendo un ottimo strumento per semplificare e riassumere i pensieri, riducendo il ragionamento umano ai suoi passi essenziali.²¹

Partendo da una riflessione sui tipi di ragionamento tradizionalmente riconosciuti dalla logica, la deduzione, l'induzione e l'abduzione, è stato Toulmin uno dei primi a sviluppare un modello che si sforzasse di rappresentare il ragionamento naturale. Lo studio in questione si concentra sull'analisi dell'argomentazione, presa in considerazione soprattutto nel suo uso quotidiano. La quotidianità, la realizzazione di processi logici nella vita della gente 'comune', è una delle caratteristiche del lavoro di Toulmin, che sottolinea l'importanza di uno sguardo teorico capace di uscire dai raffinati ma esigui ambiti interni alla materia per orientarsi verso il mondo esterno e interrogarsi sulle modalità con cui le scoperte scientifiche trovano applicazione nella vita di tutti i giorni:

²⁰ **Marvin Minsky**, *La società della mente*, Milano, Adelphi, 1989, p.366.

²¹ **Marvin Minsky**, *op.cit.*, p.360-361.

(i problemi sollevati da questi saggi) si potrebbero descrivere come problemi riguardanti la logica; sono problemi che sorgono con una forza speciale non all'interno della scienza della logica ma soltanto quando ci si ritrae per un momento dalle raffinatezze tecniche del soggetto e si indaga quale peso abbiano la scienza e le sue scoperte al loro esterno- come si applicano in pratica, e quali connessioni abbiano con i canoni e i metodi che usiamo quando, nella vita di tutti i giorni, valutiamo la correttezza, la forza e la conclusività delle argomentazioni.²²

Per Toulmin la logica è una scienza che deve occuparsi di studiare i processi di riflessione normali, corretti, razionali, evidenziando ciò che invece è da considerarsi logicamente aberrante, e quindi "patologico". È una scienza che deve studiare le abitudini di inferenza corretta e i canoni di inferenza razionali. Per realizzare tutto ciò, Toulmin fa una scelta piuttosto 'forte': rinuncia ad un approccio rigorosamente matematico, prendendo come modello la teoria del diritto.

(...)prendiamo come modello la disciplina della giurisprudenza. La logica (possiamo dire) è giurisprudenza generalizzata. Le argomentazioni possono essere paragonate ai casi giudiziari, e ciò che sosteniamo e in favore di cui argomentiamo in contesti extra-giuridici a ciò che sosteniamo nei tribunali, mentre i casi che presentiamo per convalidare ciascun tipo di affermazione possono essere paragonati tra di loro.²³

In questo modo la logica si occuperà in primo luogo di procedure, e in particolare delle procedure attraverso le quali si utilizzano determinate categorie nel "processo razionale" per difendere determinate affermazioni e confutarne altre. Le idee, più che confrontarsi rispetto ad un "ideale filosofico", in questo modo cercheranno riscontro nella "pratica effettiva della valutazione delle argomentazioni"²⁴.

L'argomentazione risulterebbe composta di cinque fattori: un dato, una

²² Stephen Toulmin, *Gli usi dell'argomentazione*, Torino, Rosenberg & Sellier, 1975, p.4.

²³ Stephen Toulmin, op.cit., p.10.

²⁴ Stephen Toulmin, op.cit., p.12.

garanzia (o regola), un fondamento, una conclusione (o tesi) e una riserva. I più importanti, di cui ovviamente non si può fare a meno sono la conclusione, i dati e la garanzia. Questa rappresentazione della fase argomentativa del ragionamento umano, mette in luce quell'elasticità e quella prudenza nel trarre delle conclusioni che è tipica di questo approccio, che evita schematizzazioni assolute e pone sempre una serie di condizioni affinché si possa accertare che un processo descritto abbia luogo. Uno degli esempi fatti da Toulmin per chiarire la sua teoria è il seguente: Petersen è uno svedese (dato); si può presumere quasi certamente che uno svedese non sia cattolico (garanzia) dal momento che la percentuale dei cattolici svedesi è minore del 2% (fondamento). Dunque, quasi certamente, Petersen non è cattolico, a meno che (riserva) entrambi i suoi genitori non siano stranieri, egli sia stato naturalizzato americano, ecc. Come ben si vede, Toulmin rinuncia alle premesse universali del tipo "tutti gli A sono B", per introdurre quelli che chiama i qualificatori (presumibilmente, probabilmente) e le ricusazioni (a meno che...). L'aspetto più interessante per i nostri scopi è dato dalla rilevazione che tale teoria scaturisce da un modello più generale del ragionamento comune, che prendiamo in considerazione. A partire da un problema, dunque, generato da un certo contesto o situazione, ha origine il ragionamento, formato da più argomenti. Ogni argomento è costituito da:

- Premesse, ovvero dei dati, degli elementi di prova, e delle regole, che permettono di inferire o meno la validità della conclusione.
- Conclusione, o tesi: il punto d'arrivo dell'argomento, verso il quale converge il processo logico, che non è quasi mai assoluta, ma deve prendere in considerazione alcune possibili riserve.

Ebbene, di fronte al sorgere di un problema, che può essere di varia natura, si adotta una strategia euristica, che, con le parole di Della Casa,

consiste nella scelta di una certa procedura di ragionamento che si ritiene adeguata alla natura del problema e nella determinazione di un orizzonte di ricerca (ossia nell'individuazione dei "luoghi" ove cercare i dati utili).²⁵

Pertanto si perviene ad una conclusione attraverso una serie di dati e regole, che spesso presentano strutture del tipo "se...allora", ma che possono

²⁵ **Maurizio Della Casa**, *Scrivere Testi*, Firenze, La Nuova Italia, 1994, p.438.

anche non venire esplicitate perché date per scontate. Naturalmente può accadere che i dati raccolti possano risultare discordanti, oppure le regole possano essere applicate solo in determinati ambiti. Ma quello che importa sottolineare è che, in questo modello logico del ragionamento comune, due sono le fasi fondamentali: la raccolta di dati, a partire dall'ipotesi (una prefigurazione congetturale della possibile conclusione) e il richiamo di regole le cui condizioni corrispondono ai dati. Solo in questo modo si potrà applicare al caso la conseguenza delle regole. Dunque, la ricerca dei dati è un passaggio fondamentale del ragionamento comune, perché permette di escludere delle ipotesi poco consistenti, di perfezionare quelle generiche o imprecise, di richiamare le regole più appropriate. Alla base del ragionamento, per questo modello, c'è la scelta di una strategia euristica efficace. Per chi si occupa di *information retrieval* questo passaggio è di importanza fondamentale, perché dimostra come siamo quotidianamente impegnati in processi di raccolta e di selezione di dati, quando commentiamo la partita al bar come quando giustifichiamo le nostre scelte politiche. Ma soprattutto, dimostra che il nostro modo di ragionare non si accontenta di una ricerca singola, ma ha bisogno di confronti, prove, riscontri di vario genere, e che alla fine di questo lavoro, non si possano escludere alcune importanti riserve. La ricerca di informazioni non è un processo rigido, formalizzabile in pochi passaggi logici. Compito di chi produce l'informazione è allora assecondare questi complessi processi logici, anziché ostacolarli proponendo una modellizzazione poco adatta al modo comune di ragionare.

L'aggiunta di qualificatori e riserve, sicuramente eleva la complessità del modello e lo rende apparentemente meno efficiente; tuttavia 'relatività' ed 'elasticità' del ragionamento sono aspetti assolutamente irrinunciabili, tanto più che, se ci riflettiamo, nella vita di tutti i giorni ci capita spesso di comportarci in maniera apparentemente illogica, cioè di violare le regole, e di trarne profitto. A tal proposito Smorti definisce "ragionamento acrobatico" il ragionamento di chi, pur essendo a conoscenza delle regole, talvolta le viola per raggiungere un accordo locale. Confrontando alcuni dati che dimostrano come nelle città nord-europee il numero di incidenti mortali sia più elevato rispetto a città i cui cittadini hanno tradizionalmente una reputazione di automobilisti indisciplinati, come Napoli, l'autore trae alcune interessanti conclusioni. Laddove, come nel nord-europa, vige una "cultura della regola", che applica una logica di tipo formale del tipo se p allora q (se il semaforo è verde, passo con tranquillità) gli incidenti sono numerosi e gravi perché la violazione di una regola da parte di qualcuno implica inevitabilmente lo

scontro. Viceversa, il ragionamento “contestuale” tipico, a detta dell’autore, dei meridionali, è così sintetizzabile: dato un contesto x , se p allora q (dato un contesto in cui tutti rispettano le regole, allora se è verde, passo. In altre parole: se è verde, passo, ma solo dopo essermi accertato che non ci sia qualcuno che contravvenendo le regole passi con il rosso). Sembra un discorso poco serio, invece è molto interessante perché dimostra come il ragionamento ‘napoletano’, che applica le regole solo dopo aver valutato il contesto, per quanto complesso e poco comodo, risulta però efficace. Smorti definisce il primo processo cognitivo un processo deterministico, limpido, trasparente, del tipo stimolo-risposta: se rosso, allora stop. Il secondo modo di ragionare, il ragionamento acrobatico appunto, è invece un modo interpretativo, più complesso, meno limpido, perché applica le regole in maniera diversa a seconda dei contesti. Per uno studio sul reperimento dell’informazione si rivela particolarmente interessante il fatto che l’autore del saggio in questione ampli le prospettive affermando che “questi due modi diversi di pensare non sono solo due processi cognitivi diversi, ma anche due diversi mondi culturali, in realtà molto collegati tra loro”²⁶. Non è qui il caso di riportare tutti i possibili mondi culturali riferibili al ragionamento tassonomico e quelli legati invece al ragionamento acrobatico, e le loro interazioni storiche²⁷; ciò che conta è rilevare che, a nostro modesto parere, i processi cognitivi che caratterizzano chi ricerca l’informazione sono sempre di tipo acrobatico, cioè implicano sempre una attività interpretativa e la capacità di adattarsi al contesto.²⁸

Il discorso sin qui portato avanti ha evidenziato l’influenza delle scienze cognitive sugli studi del linguaggio: la produzione e la ricezione dell’informazione, e quindi anche le attività di ricerca ed elaborazione, sono talvolta presentati con numerose analogie con i processi di *problem solving*²⁹. Il primo passo per la risoluzione di un problema è quello di averne una chiara rappresentazione, per poi procedere attraverso tecniche acquisite alla risoluzione dello stesso. Tale procedimento si applica al processo di ricerca dell’informazione così come a quello di scrittura: non si sta in questo

²⁶ **Andrea Smorti**, “Con il verde si passa? Il posto della violazione in un mondo di norme: il ragionamento acrobatico” in *Psicologia Contemporanea*, n°157, gennaio-febbraio 2000.

²⁷ Un bel caso di modello tassonomico rimesso in discussione da una violazione è quello dell’ornitorinco che mise in crisi gli scienziati che, avendolo scoperto, non riuscivano ad inserirlo in una delle loro categorie predefinite. Si veda a riguardo **Umberto Eco**, *Kant e l’ornitorinco*, Bompiani, Milano, 1997.

²⁸ In ogni caso, mi sento di consigliare a tutti di guardare bene prima di attraversare un incrocio anche se si ha il verde...

²⁹ **Maurizio Della Casa**, op.cit., p.26.

modo dicendo che scrivere un testo equivalga a risolvere un'equazione algebrica, ma piuttosto si sta mettendo in evidenza la necessità dell'autoconsapevolezza di una serie di fasi che bisogna attraversare per realizzare quanto ci si propone, utilizzando di volta in volta gli strumenti più adeguati. Per citare quanto scrive Della Casa,

La metacognizione è la consapevolezza che l'individuo ha dei propri processi cognitivi, e che si manifesta sia nella conoscenza delle operazioni e dei procedimenti di cui fa uso, sia nella capacità di osservarsi mentre fa un compito, ponendolo sotto controllo.³⁰

Dal *problem solving* le attività di estrazione dell'informazione possono acquisire fruttuosamente le tecniche di pianificazione e controllo dei processi. Un modello 'classico' per spiegare l'attività di scrittura come un caso di *problem solving* è quello di Hayes e Flower³¹. Dal momento che il nostro interesse primario è volto alla ricerca dell'informazione più che alla sua produzione, faremo un rapido accenno perché si tratta di un modello ampiamente sfruttato, ma non ne approfondiremo l'analisi. Secondo i due autori, sono tre le fasi determinanti un processo di scrittura: la pianificazione³², la traduzione e la revisione³³. Durante la pianificazione, il soggetto elabora gli obiettivi da realizzare attraverso il testo e stabilisce una strategia per raggiungerli tenendo conto delle informazioni di cui dispone, che gli derivano dal contesto e dalla sua memoria a lungo termine. Durante la seconda fase, la traduzione, si procede alla codificazione linguistica del materiale elaborato nella prima fase; si passa infine alla revisione, attraverso la quale il soggetto rilegge il suo elaborato ed eventualmente lo corregge. Queste tre fasi non sono disposte in sequenza rigida, ma piuttosto il soggetto

³⁰ Maurizio Della Casa, op.cit., p.31.

³¹ John R.Hayes, Linda S.Flower, "Identifying the Organisation of Writing Process", in Gregg Lee W., Steinberg E.R. (eds.), *Cognitive Process in Writing*, Erlbaum, New Jersey, Hillsdale, 1980; John R.Hayes, Linda S.Flower, "The Dynamics of Composing", in Gregg Lee W., Steinberg E.R. (eds.), op.cit.

³² La pianificazione, a sua volta, si suddivide in tre sottoprocessi: la generazione (*generating*), operazione che permette il recupero delle informazioni dalla memoria a lungo termine, l'organizzazione (*organizing*), che seleziona il materiale recuperato, e l'adattamento rispetto all'obiettivo (*goal-setting*). Quest'ultimo processo modifica il testo a seconda dell'utente finale cui si indirizza e delle caratteristiche che si vogliono dare allo scritto.

³³ Nel testo originale, i termini inglesi corrispondenti sono *planning*, *translating* e *reviewing*. Anche la revisione, come la pianificazione, è costituita da due sottofasce: la fase di *reading*, cioè di lettura vera e propria, e quella di *editing*, di correzione eventuale e perfezionamento.

le attraversa ritornando sui suoi passi o anticipando fasi successive secondo una logica circolare più che lineare. Il processo di generazione di idee, e quello di correzione, in particolare, possono interrompere gli altri processi. Pertanto il modello “specifica l’organizzazione di questi processi. In particolare, specifica un’organizzazione che è diretta ad obiettivo, ricorsiva, che consente l’interruzione dei processi e che tiene conto delle differenze individuali”³⁴. Questo modello, rapidamente presentato, ha il merito di concentrare l’attenzione sulle complesse attività di pensiero che sottostanno alla produzione di un testo, ma non sono mancate le critiche nei confronti dei suoi limiti. Della Casa³⁵ fa notare la trascuratezza con la quale vengono prese in considerazione le dimensioni linguistiche e testuali, lasciando troppo spazio alla fase di pianificazione di cui la scrittura vera e propria non sarebbe altro che una codificazione grafica. Secondo l’autore, i modelli cognitivi, enfatizzando la scrittura complessa ed adulta, dimenticano gli aspetti immaginativi, emotivi, affettivi. E tutto ciò dipenderebbe dalla “concezione computazionale della mente, vista come sistema elaboratore di informazioni, che è alla base della psicologia cognitiva”. Il discorso qui si riferisce alle attività di produzione dell’informazione, ma è generalizzabile anche a quelle di ricerca ed elaborazione. La linguistica computazionale, inutile negarlo, corre il pericolo di enfatizzare alcuni aspetti rispetto ad altri, tutte le volte che formalizzando alcuni processi mentali trascura il primo e fondamentale contesto in cui opera la mente, e cioè l’essere umano nella sua completezza.

Un’altra critica interessante alla modellistica proposta da Hayes e Flowers è quella mossa da Patrick O’Brian Holt. I modelli delle scienze cognitive, afferma l’autore, enfatizzano i meccanismi cognitivi, i flussi dell’informazione e le limitazioni o capacità del soggetto; ma falliscono soprattutto quando considerano il linguaggio come uno ‘strumento’ diretto ad un obiettivo preciso, operando una semplificazione eccessiva. Il modello di Hayes e Flowers considera i processi di elaborazione testuale come dei processi di *problem solving*, il che può andar bene, se non si esagera con le generalizzazioni; ciò che viene messo in discussione è l’utilizzo di strumenti di analisi come la *protocol analysis*³⁶, adatti ad altre attività umane. In sostanza, secondo Holt, non si può pensare ad attività testuali come la produzione testuale (e, aggiungiamo noi, il recupero dell’informazione) come

³⁴ **John R.Hayes, Linda S.Flower**, “Identifying the Organisation of Writing Process”, in Gregg Lee W., Steinberg E.R. (eds.), op.cit., p. 29, traduzione nostra.

³⁵ **Maurizio Della Casa**, op.cit., p.41.

³⁶ Tecnica che richiede all’autore del testo di verbalizzare i propri pensieri e di produrre commenti e annotazioni collaterali. Torneremo sulla *protocol analysis* nel capitolo quarto.

a processi altamente formalizzabili che hanno una sola soluzione possibile. I risultati di una ricerca informativa possono essere vari, e non ci sono prove empiriche in grado di eliminare quelli 'sbagliati'. Se si agisce attraverso l'analisi dei protocolli prodotti dagli utenti, si dà per scontato che essi sappiano perfettamente qual è l'obiettivo del loro lavoro; cosa che purtroppo non è.

A more fundamental problem arises from the lack of insight that subjects have about how they solve problems. It is not uncommon to hear them say: 'I just do it, but I'm not sure how'. Almost by definition we are not able to describe how we carry out complex tasks.³⁷

Possiamo dunque trarre delle conclusioni. Gli studi sull'*information retrieval* possono arricchirsi dell'approccio cognitivo che identifica le attività di elaborazione testuale come attività di *problem solving*: il che comporta l'importanza della metacognizione dei processi, una corretta rappresentazione dell'operazione da compiere e dei progressi effettuati. Tuttavia, occorre non cadere nell'errore di considerare questi elementi un dato di partenza: essi rappresentano piuttosto un punto d'arrivo. Chi ricerca informazione ha infatti un'idea più o meno precisa del risultato che si prefigge, ma non è in grado di descrivere il processo attraverso il quale opererà. Quando si ricerca informazione, risulta difficile rappresentare visivamente l'area concettuale in cui ci si muove, la direzione da seguire, i percorsi più brevi da compiere, perché ci sono delle ampie zone 'oscuri', dovute al fatto che stiamo esplorando, che pertanto non siamo in possesso dei dati necessari ad autorappresentare la nostra situazione.

Ciò che risulta evidente da queste riflessioni è la necessità di offrire un 'sostegno' cognitivo, al lettore come al ricercatore di testi, un sostegno che si poggia su una base di conoscenza acquisita e ben sedimentata, dalla quale si possa partire per acquisire nuova conoscenza. Colui che ricerca informazioni si muove su terreni sconosciuti, per cui il suo modo di vedere, catalogare e definire si sviluppa sulla base di conoscenze pregresse. È in tali circostanze che si rende utile il concetto di ragionamento metaforico, con le possibilità che offre ai modelli di ricerca dell'informazione. L'utilizzo del ragionamento

³⁷ Patrick O'Brian Holt, "Models of Writing: A Question of Interaction?", in Williams Noel, O'Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect, 1989, p.55.

metaforico si rivela uno strumento efficace in tutte le fasi, dalla produzione alla ricezione e fruizione dell'informazione, compresa, dunque, la fase della ricerca e dell'estrazione. Non sorprenda una tale scelta: in realtà operiamo attraverso metafore molto più spesso di quanto non ce ne rendiamo conto. Esse ci forniscono una rappresentazione di quelle 'zone oscure' della nostra conoscenza, attraverso l'utilizzo di modelli e strutture a noi più familiari. Ma prima di addentrarci in questo aspetto, sarà opportuno riflettere sull'idea di metafora, intesa proprio nella sua valenza di figura retorica, dalla quale ci sposteremo poi per i fini del nostro lavoro.

1.5. Il ragionamento metaforico e la sua rilevanza per l' information retrieval

I dispositivi alla base del funzionamento della figura retorica della metafora sono stati ampiamente approfonditi ed esiste una vasta letteratura a riguardo. Cominciamo innanzitutto a delineare una definizione 'tradizionale' di metafora; in seguito cercheremo di approfondirne il senso e l'importanza del ragionamento metaforico per la trattazione e l'elaborazione dell'informazione prendendo in considerazione in particolar modo l'analisi fatta a riguardo da Eco.³⁸ Per il dizionario, la metafora è un

Procedimento linguistico che consiste nel trasportare una parola dall'oggetto a cui la si riferisce a un altro oggetto mediante paragone sottinteso.³⁹

Dal punto di vista semiotico l'uso del termine 'parola' è troppo generico, perché non spiega quale elemento della semiosi sia coinvolto nell'operazione di 'trasporto'. Cercando una definizione 'scolastica' di metafora troviamo che essa è considerata

Il trasferimento di significato da un termine ad un altro. Nella retorica antica veniva definita una similitudine abbreviata in cui non è espresso né l'oggetto con cui si fa il paragone, né il

³⁸Umberto Eco, *Trattato di semiotica generale*, Milano, Bompiani, 1975 pp.337-352 e Umberto Eco, *I limiti dell'interpretazione*, Milano, Bompiani, 1991, pp.142-161.

³⁹Dizionario Enciclopedico Rizzoli Larousse.

termine di paragone.⁴⁰

Da questa definizione apprendiamo che l'operazione di trasporto riguarda il significato, non l'intera parola; meglio esprimersi più correttamente dicendo che essa rappresenta il trasferimento di alcune marche semantiche da un termine ad un altro. È interessante soffermarci sulla definizione della retorica antica perché essa da una parte introduce il concetto di 'similitudine', dall'altra esclude la rilevanza dell'oggetto, o meglio del referente. Questi due aspetti sono ripresi ed approfonditi da Eco il quale afferma che una metafora non solo può 'abbellire' un contenuto già dato, ma può contribuire a delineare un contenuto diverso: essa quindi può essere usata in modo creativo. La metafora e la metonimia sono figure retoriche essenziali perché rappresentano la prima un tipo di sostituzione linguistica paradigmatica, la seconda un tipo di sostituzione sintagmatica; con la prima si sostituisce per similarità, con la seconda per contiguità. La similarità che caratterizza la metafora non rappresenta una relazione tra significante e cosa significata, ma piuttosto è un caso di identità semica:

La metafora non istituisce un rapporto di similitudine tra i referenti, bensì una identità semica tra i contenuti delle espressioni, e solo mediatamente può concernere il modo in cui consideriamo i referenti⁴¹

L'interazione metaforica dunque non sostituisce referenti, e neanche espressioni, ma contenuti, considerando un contenuto in termini di componenti semantiche.

La metafora funziona perché è stato scelto, tra le proprietà periferiche di entrambi i sememi, un tratto comune che è stato eletto a genere solo ai fini di quel particolare contesto.⁴²

L'operazione di 'trasporto' di cui parlava il dizionario, dunque, riguarda solo alcune proprietà, spesso le più periferiche, di un semema, ovvero di

⁴⁰Cioffi, De Tassis, Gavazzi, Zanette, *La letteratura e le idee, dalle origini ad oggi*, Milano, Mondadori, 1989, p.1178.

⁴¹Umberto Eco, *I limiti dell'interpretazione*, Milano, Bompiani, 1991, p. 144.

⁴²Umberto Eco, op.cit., p.147.

un'unità di contenuto; le altre vengono 'narcotizzate'. Per comprendere questo concetto, utilizzeremo l'esempio fatto dallo stesso Eco, che si riferisce ad una catacresi, la forma più elementare di sostituzione metaforica: la *gamba del tavolo*. L'operazione metaforica in questo caso fornisce la funzione segnica *gamba*, costituita come tale di espressione e contenuto, ad un altro contenuto sfornito, nella lingua italiana, di espressione. Perché questo processo funzioni, non è necessaria una similitudine tra referenti, poiché una gamba umana e la gamba di una scrivania hanno forma, dimensioni, aspetto, consistenza diversi; non c'è similitudine tra espressioni linguistiche, visto che il secondo termine di paragone è completamente sprovvisto di espressione; c'è un processo di individuazione e trasporto di alcune marche semantiche, e narcotizzazione di altre. Nell'esempio in questione, si narcotizzano le marche semantiche di una gamba umana che ne fanno un oggetto vivo, mobile, costituito di tessuti organici, e si individua la marca semantica che contraddistingue la funzione di 'sostegno' della gamba, questa sì comune sia a un tavolino rococò che ai preziosi arti inferiori di Pelè.

Minsky arriva addirittura alla conclusione che ogni pensiero è in qualche misura una metafora: dal momento che non esistono due cose o due stati mentali identici, "tutti i processi psicologici devono impiegare un qualche mezzo per indurre l'illusione dell'identità", e questo mezzo è spesso una sostituzione di tipo metaforico.⁴³

L'utilizzo di 'metafore' in senso lato è stato spesso decisivo per comprendere ed analizzare alcuni fenomeni di cui non si sa molto. Il processo che avviene è quello di sostituzione di alcune aree oscure o assenti di un determinato paradigma con alcune informazioni prelevate da un altro modello utilizzato come paragone: proprio come nel caso della gamba del tavolo. Tanto per fare un esempio noto alla tradizione linguistica, si pensi allo sforzo di Saussure⁴⁴ di spiegare il gioco linguistico attraverso la metafora del gioco degli scacchi. Oppure a Chomsky, che ha costruito il suo complesso sistema di linguistica generativa utilizzando elementi della tradizione genetica. Ciò che conta in questi casi è innanzitutto testare l'efficacia della metafora, che essendo un'operazione linguistica deve rispettare alcune regole ed essere efficace, cioè produrre informazione nuova. E poi bisogna ricordare che affinché una metafora funzioni è necessario quel processo di narcotizzazione di cui parla Eco. La mancanza di questo processo di

⁴³Marvin Minsky, *La società della mente*, Milano, Adelphi, 1989, p.587.

⁴⁴Ferdinand de Saussure, *Corso di linguistica generale*, Roma-Bari, Laterza, 1992.

narcotizzazione non rende difettosa la metafora, ma ne impedisce l'esistenza stessa. La metafora di Saussure della partita a scacchi è illuminante per comprendere i concetti di sistema, valori, stati e opposizioni, ma come l'autore stesso nota, è difettosa perché il gioco degli scacchi presuppone un'intenzionalità che la lingua non ha⁴⁵. Tuttavia essa funziona, seppure nei suoi limiti e nelle sue imperfezioni, perché sono state narcotizzate marche semantiche del gioco degli scacchi quali quella di svolgersi su un'area suddivisa in sessantaquattro quadri. Se tale precisazione pare superflua è perché questo processo di narcotizzazione è talmente comune da risultare quasi automatico e da farcene dimenticare l'esistenza. La consapevolezza della sua esistenza è un contributo fondamentale che la retorica può dare agli studi di *information retrieval*. Per mezzo dell'uso di metafore è possibile infatti rendere consapevoli i lettori di processi cognitivi altrimenti di difficile comprensione. Attraverso l'uso di un linguaggio metaforico è più facile rendere rinvenibile l'informazione, poiché colui che cerca si muoverà sempre sulle basi di conoscenze pregresse ed attiverà il processo di individuazione e sostituzione di marche semantiche che è alla base del ragionamento metaforico.

Nel mondo degli elaboratori elettronici, l'utilizzo di metafore è talmente diffuso da essere ormai dato per scontato. Si pensi in particolare alla rivoluzione apportata nel mondo dei personal computer nella metà degli anni ottanta dall'apparizione delle prime GUI (Graphical User Interface). Per mezzo di esse la scrivania diventava la metafora per descrivere uno schermo, la finestra identificava il contenuto di un'applicazione, la cartella diventava metafora di un *file*. Chi non aveva mai avuto contatti con uno schermo interattivo in vita sua, aveva senz'altro già visto una scrivania di lavoro; chi non riusciva a rappresentarsi l'apertura e la chiusura di un'applicazione, trovava l'immagine della finestra pronta a riempire uno spazio semantico vuoto. Per non parlare del cestino, l'icona che, come nel caso della gamba del tavolo, fornisce un'espressione grafica ad un contenuto (cancellazione di un file) che ne era sprovvisto. Questa metafora funziona selezionando ed evidenziando la marca funzionale comune esistente tra l'icona cestino e l'oggetto cestino reale, suggerita da una analogia grafico-visiva.

Certo, ci sono anche casi in cui la narcotizzazione non si verifica come dovrebbe, e il meccanismo metaforico allora non funziona più. Si pensi alla seguente storiella, che, vera o falsa che sia, illustra bene questo aspetto:

⁴⁵ Ferdinand de Saussure, op.cit., p.109.

A Dell technician advised his customer to put his troubled floppy back in the drive and close the door. The customer asked the tech hold on, and was heard putting the phone down, getting up and crossing the room to close the door to his room.

I realizzatori di interfacce *user-friendly* conoscono bene la facilità della mente umana di memorizzare ed interpretare segni grafici più che testi scritti, e la sfruttano appieno. Il rischio che si corre è quello di superare il limite che si interpone tra una icona analogica, che permette il facile riconoscimento di una metafora, e un segno che risulta praticamente de-semantizzato se non fornito di una chiave interpretativa. Si pensi all'icona del *browser* Netscape, un faro stilizzato, che richiede un lavoro interpretativo non immediato che permetta di cogliere la metafora della navigazione marina come metafora dello scorrimento delle pagine in Internet: la maggior parte degli utenti avrà colto la metafora solo dopo aver scoperto, tramite l'esperienza, che Netscape è un *browser*. L'utilizzo di metafore è insomma molto utile, ma è un procedimento complesso che richiede sobrietà nell'applicazione e verifica dell'efficacia.

L'ultimo esempio citato ha posto in evidenza la metafora a cui siamo abituati da quando esploriamo le risorse della rete, la navigazione, talmente efficace da essere effettivamente l'unico verbo utilizzato in italiano per indicare il processo di fruizione delle pagine del *web*. Ebbene, questa metafora, pur con tutti i suoi limiti, funziona, e facilita i processi di apprendimento dell'utilizzo dei *software* di navigazione (appuntamento). Non è, tuttavia, particolarmente originale: Cadioli fa notare infatti come le metafore nautiche venissero già utilizzate dagli autori latini per indicare la composizione letteraria, e cita una poesia di Emily Dickinson che definisce il libro un "vascello veloce".⁴⁶ Ancora una volta le nuove tecnologie nascondono un'anima antica.

L'utilizzo di un linguaggio metaforico è importante perché crea consenso, mette in gioco le competenze del lettore e trasmette contenuti che possono apparire ostici, in quanto non conosciuti, facendo riferimento ai campi del sapere noti al lettore. Nell'ambito del trattamento dell'informazione, l'uso di metafore è prezioso; tuttavia, mentre il suo uso per favorire la fruizione dell'informazione è piuttosto frequente, al contrario la sua applicazione già

⁴⁶ **Alberto Cadioli**, *Il critico navigante [saggio sull'ipertesto e la critica letteraria]*, Genova, Marietti, 1998, p.9.

nei processi produttivi è più rara.

Il ragionare utilizzando metafore, si è detto, permette di operare in aree poco conosciute del sapere facendo riferimento a modelli analoghi ben analizzati. L'osservazione di situazioni complesse realizzata supportandosi con modelli metaforici facilita la comprensione di processi altrimenti invisibili. È ovvio che non si potrà impiantare completamente in un'area di studi un modello proveniente da un'altra disciplina, ma solo alcuni aspetti ritenuti rilevanti, e solo se questi ultimi sono stati effettivamente studiati a fondo; insomma, il processo di narcotizzazione è sempre indispensabile. Inoltre, più metafore collegate funzioneranno meglio che una soltanto, perché l'una può cogliere un aspetto che l'altra trascura. Dati questi presupposti, le prospettive, anche per la ricerca dell'informazione, sono importanti:

Accurate rethinking of the inner nature of information and consistent redefinition of communicative interactions by introducing physical models of reasoning and by proposing and building metaphorical environments for taking action are in fact fundamental steps toward better coordination of any information-seeking.⁴⁷

Prendendo in prestito modelli e sistemi da altre scienze di provata tradizione come la fisica, la chimica, la meccanica, è possibile, nei limiti permessi dal processo analogico, comprendere più a fondo i processi che caratterizzano la comunicazione e l'informazione. La metafora diviene pertanto un efficace strumento di interpretazione e comprensione. In particolare è possibile rendere intelligibili e più facilmente comprensibili un'insieme di processi cognitivi attivati da chi produce informazione. Tale principio è alla base degli sviluppi di una metodologia per il trattamento dell'informazione, il modello CPP-TRS, ideata e realizzata da Graziella Tonfoni⁴⁸.

Tale metodologia mette in diretta relazione attività come la progettazione, il design, la pianificazione, solitamente collegate all'ambito industriale, con

⁴⁷ **Graziella Tonfoni**, *Information Design. The Knowledge Architect's Toolkit*, Scarecrow Press, Lanham, Maryland, 1998, p.4.

⁴⁸ I testi di riferimento per l'approfondimento della metodologia sono i seguenti: **Graziella Tonfoni**, *Scrittura multimediale*, Treviso, Pagus, 1991; *La comunicazione aziendale come arte visiva*, Treviso, Pagus, 1991; *Partitura, solfeggio, movimento*, Treviso, Pagus, 1992; *Abitare il testo*, Treviso, Pagus, 1995.

la scrittura e l'elaborazione del testo, sulla scia degli studi condotti dalle scienze cognitive. L'utilizzo di tecniche di progettazione, di elaborazione ma anche di estrazione testuale si mostra tanto più utile quanto più si considerino le prospettive che le nuove tecnologie offrono all'ambito comunicativo.

Prima di occuparci di questo modello, che conclude questo capitolo iniziale della nostra indagine, nel quale si è cercato di fornire le basi culturali e gli strumenti cognitivi che costituiscono le fondazioni teoriche dell'intera ricerca, sarà necessario dedicare un po' di spazio all'area di studi in cui si inserisce. Si tratta dell'*information design*, un settore di ricerche che si occupa dell'elaborazione dell'informazione. Si è già ribadita l'idea per cui tanto più accurata è l'elaborazione dell'informazione, tanto più facilitato sarà poi il suo recupero. L'*information design*, dunque, occupa una posizione a monte, in un'immaginaria mappa dei processi di elaborazione del testo, rispetto all'*information retrieval*, e di conseguenza non può che influenzarne pesantemente gli sviluppi e le possibilità.

1.6. L'Information Design⁴⁹

Il seguente breve *excursus* nel mondo dell'*information design* richiederebbe maggiori approfondimenti e annotazioni, vista la complessità e la dinamicità del campo di studi. Gli elementi forniti sono appena sufficienti per comprendere quanto alto sia il livello di generalità della materia e quanto importante sia la capacità di discriminazione. Sarà infatti solo distinguendo con accuratezza ciò che è pertinente da ciò che non lo è, nell'*information design*, che si eviterà il rischio serio che la complessità si traduca in genericità.

L'*Information Design* è un'area di studi che si occupa dell'elaborazione quantitativa e qualitativa dell'informazione: oggetto di studio sono modelli diversificati di elaborazione dell'informazione, con applicazioni che spaziano tra aree tra loro anche piuttosto distanti; si pensi all'area della comunicazione aziendale, all'elaborazione di manuali di istruzione, o di controllo di sistemi di volo.

Per comprendere a fondo l'origine dell'*information design*, si consideri il significato dei due termini: per la parola '*information*', è necessario risalire alle radici latine. Informare, infatti, significa dare forma. L'*information*

⁴⁹ Cfr. **Graziella Tonfoni**, *Linguistica Computazionale. Appunti del corso 1998-99*, Università di Bologna; *Information Design. The Knowledge Architect's Toolkit*, Lanham, Maryland, Scarecrow Press, 1998.

design, in quest'ottica, può definirsi l'arte e la tecnica di organizzare e impacchettare le notizie e i dati affinché diventino informazione. Per comprendere il termine *design*, invece, si tenga presente il ruolo che durante la rivoluzione industriale ha ricoperto la fase del processo produttivo costituita dalle attività di pianificazione e di progettazione: mentre l'artigiano progetta l'oggetto e ne segue tutte le fasi di produzione sino alla realizzazione finale, il *designer* si limita alla progettazione iniziale, cui seguirà una produzione in serie effettuata da altri. La metafora spaziale è fondamentale per comprendere a pieno questa disciplina, come si intuisce da questo passo estratto da Tonfoni:

An information designer needs to be able to plan consistently and to organize knowledge buildings as well as information territories and open spaces (...) Applying metaphors to information design means creating a conceptual platform for supporting newly defined concepts with derivative experience coming from other fields that are well known.⁵⁰

L'*information design* si caratterizza per la presenza di figure diverse: ingegneri, scrittori tecnici, artisti grafici, esperti di elaborazione dati. Questa ricchezza di specializzazioni ha portato ad un cambiamento di prospettiva, poiché l'anima più 'creativa' si è aggiunta alla tradizionale anima ingegneristico-costruttiva. La formazione dei *designer* è necessariamente complessa, richiede scientificità, accuratezza e creatività, e la capacità di gestire un'area di conoscenza molto variegata.

Occorre a questo punto chiarire un aspetto: l'oggetto di trattazione dell'*information designer* non è il testo nella sua accezione più estesa; il documento che il *designer* prende in considerazione è un oggetto informativo la cui caratteristica sostanziale è quella di trasferire informazione filtrata ed attendibile. Non si comprendono pertanto tra gli oggetti di analisi i testi letterari, che seguono logiche interne non direttamente riducibili alle funzioni di esaustività, coerenza, efficacia, brevità, che devono caratterizzare, invece, una trattazione manualistica.

Un altro concetto chiave nell'*information design* è quello di *flusso informativo testuale*: ovvero informazione strutturata in un processo di

⁵⁰ **Graziella Tonfoni**, *Information Design. The Knowledge Architect's Toolkit*, Scarecrow Press, Lanham, Maryland, 1998, p.2-4.

elaborazione continua che produce come risultato un testo documentale. La continuità di tale processo fa sì che il testo documentale si articoli e si modifichi nel tempo senza raggiungere mai una situazione di assoluta staticità. La 'documentazione' che genera il flusso testuale, infatti, è la fotografia di un divenire ed è sottoponibile a continue revisioni e modifiche, generando come conseguenza una continua instabilità. Quello dell'*information design* è dunque un dominio instabile.

Da quanto detto sinora emerge l'importanza per l'*information designer* di una fase di pianificazione preventiva. In particolare, si parla di fase predocumentale in riferimento al processo di ricerca e di identificazione dei vari pacchetti informativi che andranno poi a costituire il documento. Questa fase di pianificazione comprende l'estimo relativo agli investimenti necessari, nonché ai tempi di produzione, agli strumenti da adottare, alle competenze richieste, eccetera, nella quale si accerta quantitativamente che tutti gli elementi importanti siano trattati adeguatamente.

Per quanto possano essere accurate la pianificazione e la fase predocumentale, non sempre è possibile programmare le fasi di discontinuità: ci possono essere momenti di turbolenza interna, causati da problemi non preventivati. È pertanto fondamentale specificare subito a quale utente ci si rivolge, affinché il testo sia quantitativamente corretto e qualitativamente pertinente. Bisogna cioè conoscere la 'direzione' del flusso testuale, anche perché i modelli d'utente coinvolti dallo stesso flusso possono essere più di uno. Il controllo dell'informazione va fatto prima e non dopo, per risparmiare all'utente tempo ed energia, dal momento che l'organizzazione del flusso informativo è spesso più un 'togliere' che un 'aggiungere'. Per ciò che concerne l'informazione, infatti, il concetto di sovrabbondanza non è affatto garanzia di qualità. Ne consegue quanto già affermato in precedenza: per i testi esplicativi e informativi la staticità e la finitezza sono piuttosto rare. L'informazione è sempre da relativizzarsi rispetto alle esigenze dell'utente: quanto è rilevante, non sempre è pertinente.

Negli Stati Uniti ha avuto luogo recentemente una riformulazione dell'*information design* che ha prodotto l'esigenza del nuovo termine di *knowledge architecture*, ovvero architettura della conoscenza. Secondo gli appartenenti a questa corrente, che amano definirsi '*knowledge architects*', l'*information design* si iscrive nel solco della tradizione ingegneristica, mentre altri ne sottolineano piuttosto l'aspetto estetico ed artistico. L'architetto della conoscenza è responsabile di una pianificazione

incrementale: mentre il *designer* pianifica per la produzione, l'architetto prevede la dinamica costruttiva del suo prodotto, la sua evoluzione. La pianificazione iniziale, originaria, è riadattata continuamente, non si conclude mai definitivamente. L'architetto media stili e conoscenze, accompagna le conoscenze tecniche alle sue scelte stilistiche, è interessato all'utente e alle dinamiche che l'oggetto informativo instaura.

Come vedremo in seguito, parlando del ruolo dell'autore di un ipertesto, lo scrittore che oggi faccia uso di nuove tecnologie finisce inevitabilmente per divenire un 'architetto' della conoscenza. Zimmerman⁵¹ suggerisce che la professione dello scrittore, che essendo una costruzione sociale dipende dall'evoluzione della società, è chiamata oggi a ripensarsi completamente per adattarsi all'ambiente fornito dai nuovi strumenti di comunicazione. Lo scrittore ipertestuale diventa un architetto dell'informazione in grado di costruire spazi informativi di n-dimensioni, e non può opporsi alla "ricostruzione della sua professione".

1.7. La metodologia CPP-TRS

La metodologia CPP-TRS (Communicative Program Positioning- Text Representation System) si pone come obiettivo l'aumento della conoscenza e il miglioramento dell'uso dei vari codici linguistici che caratterizzano gli atteggiamenti comunicativi. Per far ciò è indispensabile la presa di consapevolezza delle potenzialità di questi codici, della loro possibilità di interazione.

Gli obiettivi del CPP-TRS sono: A) Accrescere la capacità di riconoscere e classificare le varie tipologie testuali e comunicative sulla base di una analisi e riconoscimento delle rispettive proprietà e caratteristiche. B) Sviluppare la capacità di produrre-applicare criteri e schemi (verificandone volta per volta la coerenza rispetto ai modelli) per un'organizzazione delle forme testuali nello spazio. C) Educare ad una lettura ed interpretazione critica dei testi già prodotti per aumentare la sensibilità percettiva e valutativa nei confronti di quanto viene progettato in vista di obiettivi volta per volta

⁵¹ **Muriel Zimmerman**, "Reconstruction of a Profession: New Roles for Writers in the Computer Industry", in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press, 1989.

predisposti. D)Sviluppare una percezione testuale globale e complessa che tenga presenti le varie componenti rilevanti, in particolare visive, spaziali,concettuali, logiche.) Sviluppare una sensibilità relativa alle specificità dei codici e dei canali e dei mutamenti che possono subentrare nel caso dei processi di produzione testuale e di trasferimento e passaggio da un codice ad un altro, nonché dei processi di interpretazione.⁵²

La metodologia si prefigge il raggiungimento di tali obiettivi attraverso l'utilizzo di una semantica visiva, cioè di una serie di metafore visive che permettono, come già sostenuto in questa ricerca, una più immediata descrizione di fenomeni altrimenti incomprensibili, e quindi il concepimento di modi diversi di rappresentazione della conoscenza, di organizzazione di momenti comunicativi. Alla base di questa metodologia c'è il proposito di "rappresentare aspetti invisibili del linguaggio naturale trasformandoli in elementi tangibili e visibili"⁵³. Poiché il panorama attuale presenta infatti un aumento impressionante dei canali comunicativi, diventa essenziale per l'individuo il miglioramento della capacità personale di conoscenza e gestione dei codici, a seconda delle proprie tendenze e necessità. Grazie a tale capacità il soggetto viene posto in grado di selezionare gli stili e gli atteggiamenti comunicativi adatti alle situazioni che si trova a gestire. Scrive Tonfoni:

L'idea base è che il testo è una scrittura dinamica che si viene costruendo: di conseguenza la "costruzione del testo" è vista come una operazione di conoscenza, di approfondimento e di presa di coscienza delle varie fasi di tale progetto. Lo scrittore, chiamato ad affrontare problemi di progettazione, si avvicinerà alla costruzione del testo avendo in mente obiettivi ben definiti: sarà quindi una logica progettuale a determinare il tipo specifico di coerenza del testo in relazione ai suoi obiettivi

⁵² Cfr. **Graziella Tonfoni**, *Abitare il testo*, Treviso, Pagus, 1995, p.13.

⁵³ **Graziella Tonfoni**, *Communication Patterns and Textual Forms*, Exeter, Intellect, 1996, p.7, traduzione nostra.

comunicativi.⁵⁴

Noi ci occuperemo soltanto, per ovvie necessità di delimitazione dell'area, della parte della metodologia che suggerisce l'utilizzo di strumenti per l'attuazione di processi di *information retrieval*. D'altronde, tale metodologia trova applicazione tutte le volte che si desidera alzare il livello percettivo-visuale della comunicazione: si rinforza il messaggio con degli strumenti che, da una parte chiariscono ciò che si intende dire, dall'altra favoriscono la comprensione e il successivo reperimento dell'informazione. In particolare, nella nostra attività si sono rivelati particolarmente utili gli strumenti definiti da Tonfoni *Segnali*. Il segnale è uno "stimolo visivo che anticipa e fa scattare una azione o reazione comunicativa specifica", "che attiva una reazione percettiva di interpretazione immediata". Tali segnali testuali producono una percezione visiva che è produttiva di contenuto, anche se ad un livello pre-linguistico. Attraverso la loro presenza, si suggeriscono al lettore delle aspettative in relazione al testo prima ancora che questo appaia; e, una volta che anche la dimensione linguistica è stata pienamente fruita, tali segnali restano come potenti strumenti per il recupero successivo dell'informazione. I segnali proposti dal CPP-TRS sono costituiti da chiavi di lettura testuale: lo scrittore indica al lettore che il suo testo è narrativo, è un riassunto, un ricordo, un concetto astratto, un commento, un'analogia, un collegamento, un'idea globale, un'espansione, un'alternativa. Nell'ambito linguistico, ci sono alcuni elementi visivi che ci forniscono un'immediata interpretazione linguistica del periodo: si pensi al punto interrogativo, e alla potenza semantica di questo semplice carattere grafico. Ebbene, i segnali permettono di fornire questo genere di informazioni sul processo cognitivo che l'autore ha effettuato per realizzare un testo. Perciò richiedono, da parte dello scrittore, la piena consapevolezza del testo gestito. Essi al tempo stesso forniscono una mappa visiva per il lettore che percepisce meglio la struttura del testo che sta analizzando. Qui non è nostra intenzione approfondire la valenza dei segnali della metodologia, perché sarebbe un lavoro che ci

⁵⁴ **Graziella Tonfoni**, *Scrittura multimediale*, Treviso, Pagus, 1991, p. 17; **Graziella Tonfoni**, "Intelligent control and monitoring of strategic documentation: a complex system for knowledge miniaturization and text iconization", in *The Proceedings of the ISIC/CIRA/ISAS 98 Conference*, Gaithersburg, National Institute of Standards and Technology, U.S. Dept. of Commerce, MD, 1998; **Graziella Tonfoni**, "On augmenting documentation reliability through communicative context transport" in *The Proceedings of the 1999 Symposium on Document Image Understanding Technology*, Annapolis, MD, 1999.

allontanerebbe dai nostri obiettivi. Quel che importa rilevare è che la presenza di un segnale grafico chiarisce al lettore la dinamicità del processo di lettura, che può essere direzionato e subire cambi di ritmo; fornisce un surplus semantico sul testo nella sua completezza, favorisce il recupero dell'informazione. Tale affermazione potrebbe apparire ovvia, se si pensa alla funzione paratestuale di titoli, capitoli, paragrafi; ma risulterà meno banale quando si penserà ai suoi possibili sviluppi nel mondo dell'informazione elettronica, come vedremo nel quarto capitolo, quando torneremo ad occuparci dei segnali trattando i sistemi di codifica elettronica solitamente definiti *mark-up*. Per il momento sarà sufficiente fornire un'anticipazione sul lavoro analizzato nella seconda parte della ricerca. Nell'elaborare un testo che contenesse continui rimandi intertestuali, si è fatto largo uso di segnali grafici volti a fornire degli strumenti efficaci di recupero dell'informazione, proprio nell'ottica di quanto affermato in queste pagine. Di fronte ad un'icona stilizzata posta al principio di una frase, il lettore è stato posto in grado di comprendere immediatamente la provenienza della citazione, ed è stato dotato pertanto di una chiave interpretativa per la lettura. Non solo: in questo modo egli può rapidamente evitare i brani tratti da un testo, se lo ritiene poco importante per il suo lavoro, e recuperare esclusivamente quelli che reputa più interessanti, semplicemente osservando le icone poste a margine. L'argomento sarà approfondito ampiamente in seguito, per cui non è il caso di dilungarci oltre.

Dopo aver sin qui prospettato la base scientifico-culturale da cui muove questo lavoro, nel prossimo capitolo si cercherà di analizzare, nei limiti adeguati a questo studio, il rapporto tra elaboratore elettronico e testualità, per comprendere meglio le possibilità e i problemi che il computer ha portato ai ricercatori di informazione.

2. Le nuove tecnologie per l'elaborazione testuale

"Devo comprare una nuova macchina da scrivere. Questa che ho fa troppi errori!"
(L. Fetchner)

2.1. La scrittura con il computer

Gli universi originariamente distanti del computer e del testo linguistico hanno subito nel corso degli anni un processo di graduale avvicinamento, fino a dare origine a veri e propri nuovi mezzi di comunicazione che sono solitamente raggruppati sotto il termine 'ombrello' di *new media*, o ipermedia.

La chiave di volta per comprendere l'importanza dell'interazione tra macchina e utente, che poniamo come premessa indispensabile a qualsiasi analisi o riflessione successiva, è data dal principio che l'uomo è in grado di svolgere meglio alcune attività, la macchina altre: pertanto l'adozione di sistemi elettronici per la fruizione (produzione, elaborazione, ricerca) testuale dovrà sempre muoversi tenendo presente le caratteristiche dell'uno e dell'altro 'partner'. Brown⁵⁵ parla di liste "MABA-MABA", con un acrostico che sta per "Men Are Better At..." "Machines Are Better At...". Queste liste non sono definitive, poiché gli sviluppi della ricerca nell'ambito dell'Intelligenza Artificiale migliorano continuamente le possibilità dei computer, così come gli studi di psicologia cognitiva forniscono nuovi spunti da prendere in considerazione a riguardo della mente umana. Tra le differenze di cui prendere nota, riportiamo nella seguente tabella quelle che riteniamo più rilevanti:⁵⁶

⁵⁵ Marlin C. Brown, *Human-Computer Interaction Design Guidelines*, Exeter, Intellect, 1999 p.5.

⁵⁶ La tabella è il risultato di una rielaborazione, attraverso la quale si sono eliminati alcuni elementi e se ne sono aggiunti altri, di una tabella affine presentata da Marlin C. Brown, op.cit., p.6.

Gli esseri umani fanno meglio:

Avvertire stimoli in un contesto 'rumoroso'
Riconoscere modelli simili in situazioni differenti
Ricordare principi e strategie
Imparare dall'esperienza
Agire in condizioni inaspettate e adattare le decisioni alla situazione
Concentrarsi su una specifica attività
Applicare nuovi principi per risolvere vari problemi

I computer fanno meglio:

Avvertire stimoli al di là della portata umana, ma solo in base a modelli predefiniti
Misurare e contare quantità fisiche
Fornire risposte immediate in base a stimoli ricevuti
Ragionare deduttivamente; inferire conseguenze solo da principi generali ben definiti
Memorizzare accuratamente grosse quantità di informazione
Compiere numerose attività contemporaneamente
Lavorare in buone condizioni per periodi estesi di tempo e con carichi notevoli di informazioni

Tabella 1: confronto tra le competenze umane e quelle del computer

Sono solo alcuni degli aspetti che possono caratterizzare le differenze tra mente umana e intelligenza artificiale. In sostanza, da questa tabella possiamo trarre le seguenti conclusioni valide negli ambiti di applicazione dell'*information retrieval*: l'uomo, al contrario del calcolatore, è in grado di memorizzare poco e in maniera poco affidabile; ma è in grado di sviluppare nuove strategie e nuovi metodi per risolvere i problemi. Il calcolatore svolge i suoi compiti in maniera perfetta, non avverte fatica e può essere rapidissimo; ma le sue capacità si limitano a ciò per cui è stato programmato: non è in grado di affrontare situazioni impreviste, né di imparare dall'esperienza. Insomma, il calcolatore dà il meglio di sé come *protesi* della mente umana: è in grado di amplificarne enormemente le possibilità, ma necessita della guida dell'individuo. Questi principi, apparentemente scontati, sono invece indispensabili per comprendere le linee di direzione che dovrebbe intraprendere lo sviluppo di *software* per l'*information retrieval*. Quest'ultimo si rivela poco efficiente tutte le volte che costringe l'uomo a memorizzare molta informazione (per esempio, un programma che

richiede l'apprendimento di lunghe stringhe di sintassi in linguaggio macchina), quando gli impedisce di intraprendere percorsi di ricerca alternativi e non pianificati dalla macchina (per esempio, gli ipertesti che costringono in un percorso di collegamenti il lettore, impedendogli di saltare le parti meno interessanti), quando ancora richiede al computer di svolgere attività di 'interpretazione' di cui la macchina non è capace: il computer, infatti, per quanto complesso può essere stato il lavoro di programmazione, adotta comunque un comportamento comunicativo del tipo 'stimolo-risposta', non essendo capace dell'invenzione necessaria all'interpretazione. Un ultimo aspetto differenzia l'operatore umano dal computer. L'abbiamo lasciato in fondo alla lista perché è quello di maggiore importanza per il recupero dell'informazione: il computer, attuando una ricerca, è in grado di prendere in considerazione esclusivamente il piano dell'espressione della funzione segnica, mentre l'essere umano è in grado di analizzarne anche il piano contenutistico, disambiguandolo in base ad un contesto pragmatico. Su questo aspetto di fondamentale importanza ritorneremo in seguito. Abbiamo infatti inserito in precedenza, tra le basi della nostra ricerca, la linguistica testuale, proprio per sottolineare il valore che attribuiamo al testo come unità di contenuto e non solo unità espressiva: purtroppo oggi l'aspetto semantico e pragmatico del testo non sono analizzabili con il calcolatore elettronico, nonostante gli sforzi compiuti dai ricercatori di Intelligenza Artificiale non escludano che una tale possibilità si realizzi in futuro.

In effetti il computer, nel corso di qualche decennio, ha cambiato completamente le sue prospettive di sviluppo e di progettazione, indipendentemente dalla volontà dei realizzatori. Infatti l'enorme diffusione dei *personal computer*, dovuta a fattori economici, industriali e sociali oltre che tecnologici, ha generato un imprevisto cambio di rotta nel cammino di ricerca: da strumento in grado di fornire nuove possibilità e nuovi orizzonti creativi per lo scrittore e il lettore, esso si è trasformato in macchina che non aggiunge alcuna competenza, ma permette di svolgere le stesse attività svolte in precedenza, ma più velocemente:

Since the introduction of inexpensive and powerful personal computers, we have seen a change in focus away from developing such new strategies toward finding ways to do the old things faster(...) Instead of enabling scholars to perform tasks that were not possible before, today's

system emulate typewriters.⁵⁷

Da principio analizzeremo questa seconda tendenza, cioè il modo in cui l'elaboratore elettronico si è avvicinato al mondo della testualità come semplice 'attrezzo' di scrittura. Il computer insomma come una macchina da scrivere, o poco più; ed è altamente probabile che tuttora per la maggior parte degli utenti il *personal computer* non sia altro che una sofisticata scrivente, un comodo strumento per la scrittura lineare 'tradizionale', destinata cioè al supporto cartaceo: relazioni, lettere, e documenti che comunque non sfruttano le potenzialità multimediali dell'elaboratore. Quest'utilizzo così diffuso merita una seppur breve riflessione, perché ha apportato delle modifiche alle modalità di produzione dell'informazione, che hanno avuto non pochi riflessi sulle modalità successive di memorizzazione e ricerca della stessa.

2.1.2. Il computer come macchina da scrivere

Abbiamo in precedenza evidenziato come gli esperti di *information design* che hanno sviluppato le interfacce tra utente e macchina abbiano attinto a piene mani dall'universo di conoscenze precedenti, per evitare che il carico cognitivo necessario all'apprendimento delle tecniche di utilizzo di un nuovo strumento potesse scoraggiare la sua diffusione.

In human-computer interface design, the "computer as a tool" philosophy implies that designers must actively pursue techniques to reduce the mental processing operations required just to be able to use the tool. Mental processing operations include requirements for the user to learn complex commands and syntax, memorized encrypted codes and abbreviations, or translate data into other units or formats before they can be applied to the problem at hand.⁵⁸

Da qui, l'enorme valenza dell'utilizzo di interfacce grafiche (GUI) per l'apprendimento dell'uso del calcolatore attraverso l'iconismo e il linguaggio metaforico. Ma questa enorme facilitazione nell'uso del computer come

⁵⁷ **James H. Coombs, Allen H. Renear, Steven J. DeRose**, "Markup Systems and the Future of Scholarly Text processing", in *Communications of the ACM*, V.30, n°11, Novembre 1987, p.933.

⁵⁸ **Marlin C. Brown**, *Human-Computer Interaction Design Guidelines*, Exeter, Intellect, 1999, p.4.

strumento di scrittura ha causato degli effetti non previsti, o, come si dice nella teoria dei mass-media, “effetti perversi”. È infatti ormai opinione condivisa che l'enorme diffusione dell'utilizzo del *personal computer* per l'elaborazione dei testi abbia avuto una certa influenza sul processo di scrittura in sé. Dapprima utilizzato solo a livello professionale (nelle case editrici, nelle redazioni giornalistiche) e solo nell'ultima fase della produzione testuale, quella cioè di revisione e realizzazione grafica (*editing*), oggi il computer è usato un po' da tutti e già durante le prime fasi della scrittura. Questa stessa tesi è stata realizzata utilizzando un programma elettronico di scrittura (*word-processor*) e non ha un lavoro manoscritto 'precedente' alle spalle. I vantaggi, ma anche i pericoli di un cattivo uso, sono parecchi, come evidenzia Serafini⁵⁹. Un testo può essere modificato, cancellato, spostato, senza nessuna fatica fisica, e può essere elaborato e stampato rendendolo un prodotto finito che non ha molto da invidiare, almeno a livello estetico, ad un testo editoriale. Tuttavia questa 'bellezza' ha delle controindicazioni: il testo che appare sul video o stampato può essere considerato finito proprio per la sua bellezza esteriore, quando invece necessiterebbe di lavoro e perfezionamento nei contenuti.

Tra gli utenti di WP è abbastanza diffuso l'errore di cominciare troppo presto la produzione del testo, confidando nella magia della macchina e saltando tutto il lavoro preparatorio. Si deve tener presente che, anche se la modifica locale di un testo tramite WP è facile, il cambiamento dell'impostazione è difficile o impossibile ("raddrizzare" un testo richiede talvolta più energie che riscriverlo completamente). Questo modo di procedere, tipico dei neofiti del WP, è pericoloso ed ingenuo perché veramente dispersivo. In alcune facoltà universitarie i docenti sostengono che la qualità delle tesi di laurea è calata da quando i ragazzi hanno iniziato ad usare i WP.⁶⁰

Serafini afferma che il rischio più grosso che corre chi utilizzi in maniera inappropriata un elaboratore elettronico è quello di adottare una strategia di generazione testi *bottom-up* laddove sarebbe preferibile un approccio *top-*

⁵⁹ MariaTeresa Serafini, *Come si scrive*, Milano, Bompiani, 1992, pp.339-347.

⁶⁰ Maria Teresa Serafini, op.cit., p.345.

down. Nel primo caso, si procede ad una produzione testuale che, attraverso l'accumulo di elementi anche eterogenei, procede senza un criterio definito da principio, ed opera una serie di interventi successivi di correzione ed aggiustamento: è insomma una produzione del testo graduale che procede un po' per tentativi. Il secondo approccio, invece, stabilisce sin dall'inizio una struttura ed una organizzazione nella quale inserire successivamente le informazioni. La scaletta di partenza è così ricca che spesso non sono richieste ulteriori notevoli riorganizzazioni.

Inoltre i testi elaborati con il computer tendono ad essere più schematici ed ordinati, ma anche meno dotati di connettivi e collegamenti tra i vari paragrafi che sono presentati 'autonomamente', visto che il processo di scrittura può procedere a balzi e non necessariamente svilupparsi linearmente. I testi prodotti con l'ausilio dell'elaboratore elettronico risulterebbero anche più ridondanti e verbosi, perché lo scrittore farebbe troppo affidamento nella fase di revisione, in assoluto la più favorita dall'utilizzo del computer. In breve, se scrivere con il computer diventa più facile, diventa più facile scrivere male. L'autrice, è opportuno sottolinearlo, non riporta alcun dato statistico a sostegno delle sue tesi.

In ogni caso, è altamente probabile che il computer, anziché proporsi come strumento in grado di avvantaggiare lo scrittore fornendogli degli strumenti di metacognizione, finisca talvolta per 'offuscare' ancora di più la struttura immanente ad un testo, concentrandosi esclusivamente sulla sua forma più superficiale. La facilità della produzione testuale con il computer rischia di portare a trascurare i processi di generazione del testo 'prelinguistici', quelle fasi di pianificazione del processo di scrittura che invece sono alla base delle teorie dell'*information design*. Tutto ciò complica notevolmente il lavoro successivo di recupero dell'informazione, dal momento che si rischia di esasperare la prolissità del testo a discapito della sua sinteticità.

Non è il caso di trarre giudizi troppo assoluti o generici: come ci ricorda Landow⁶¹, è vero che l'utilizzo della tecnologia digitale può 'falsificare' il testo, o meglio trasformarlo sotto certi aspetti; tuttavia questo processo di influenza della tecnologia sulla scrittura, sulla cui importanza non tutti gli autori concordano⁶², si è senz'altro verificato anche prima della nascita del

⁶¹ **George P. Landow**, "The Convergence of Contemporary Critical Theory and Tecnology", in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.87.

⁶² Il dibattito a riguardo è davvero vasto: semplificando drasticamente, possiamo dire che c'è una tradizione statunitense, che fa capo soprattutto ai lavori di Marshall McLuhan, e che ha tra i suoi fautori

calcolatore. Anche la tecnologia della stampa ha causato non poche trasformazioni nella cultura medievale, così come la pagina scritta aveva ‘sconvolto’ il discorso orale.

Vediamo adesso come il computer possa essere utilizzato anche come strumento ‘intelligente’, aiutando l’autore nello svolgimento dei suoi compiti cognitivi, e non solo in quelli manuali, e apportando così nuove possibilità di ‘trasformazione’ al processo di scrittura.

2.1.3. Il computer come ausilio intelligente alla scrittura

Il computer, anche nel momento in cui si pone come ausilio alla fase creativa della scrittura, non può trascurare innanzitutto le esigenze dell’utente. Per cui, alla base di un buon programma di scrittura, sostiene Williams⁶³, deve esserci la conoscenza di queste esigenze. Il programma illustrato, il Ruskin, è progettato per essere quanto più flessibile possibile, fornisce pieno controllo all’utente e un’interfaccia che si sforza di essere gradevole. In effetti le esigenze di uno scrittore accademico che voglia migliorare i propri scritti sono differenti da quelle di una segretaria, che agisce sul testo prodotto da qualcun altro, come da quelle di uno studente che ha bisogno di imparare a produrre testi e non solo di perfezionare lo stile.

Il programma in questione fornisce supporto sia per la scrittura vera e propria che per l’apprendimento delle modalità di scrittura, ma solo nella fase cosiddetta *postwriting*, partendo cioè da un testo già elaborato. Sulla base di una serie di regole di ‘buona scrittura’ che i programmatori hanno inserito, il computer dà dei consigli sulla lunghezza dei periodi, sugli errori di ortografia, sulla sintassi utilizzata, sul vocabolario. La flessibilità del programma e il controllo da parte dell’utente sono garantiti dal fatto che quest’ultimo può impostare il tipo di analisi che l’elaboratore deve attuare sul testo. Il ragionamento che permette all’elaboratore di fornire consigli è del seguente tipo: se il livello di istruzione del destinatario del testo è basso, allora le frasi complesse dovrebbero essere numericamente limitate; se il testo è lungo, e il

più convinti Elizabeth Eisenstein, Ted Nelson, e, tra i più moderati, Jay David Bolter e George P. Landow, per la quale l’influenza della tecnologia sulla scrittura e sulla cultura in genere è talmente elevata da determinare dei veri e propri cambi di paradigma culturale. Contrapposta a questa c’è una tradizione, con origini e diffusione soprattutto europee, che ridimensiona l’importanza della tecnologia come fautrice di cambiamento culturale, rivalutando le variabili sociali, economiche, storiche. Come dovrebbe ormai essere apparso chiaro, questa seconda prospettiva è quella che ha influenzato maggiormente questo lavoro, in cui tuttavia si è sempre cercato di evitare posizioni estremistiche.

⁶³ Noel Williams, “Computer assisted writing software: Ruskin”, in Williams Noel, O’Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect, 1989, p.1.

testo è complesso, e la relazione tra emittente e destinatario è formale, allora è consigliabile usare dei titoli intermedi. In pratica, il lavoro svolto dal computer è quello di riconoscere occorrenze che si adattano a modelli predefiniti in memoria, selezionando di conseguenza i consigli da fornire in base ad una sua libreria interna. La logica interpretativa del computer è sempre quella stimolo-risposta (se...allora): essa funziona bene in presenza di regole abbastanza rigide (come quelle della sintassi), un po' meno in caso di regole meno rigide, come quelle della scelta lessicale. Moltissimi utenti dei calcolatori trovano molto comodo far operare al computer la correzione ortografica del testo; molti meno sono gli utenti disposti ad accettare i consigli di sostituzione lessicale di termini. Finché l'elaboratore opererà in base a modelli predefiniti, non potrà operare un'attività di 'consigliere' tutte quelle volte che il testo non corrisponderà ad uno di questi modelli; e per quanto vasta possa essere la memoria dell'elaboratore, esisterà sempre un caso che non è stato preso in considerazione dal programmatore. Pertanto, questo genere di ausilio sarà davvero 'intelligente' solo quando sarà in grado di produrre nuovi modelli, cioè di apprendere dall'esperienza; è la strada sulla quale sono indirizzati gli studi sui Sistemi Esperti, per i quali comunque c'è ancora molto lavoro da fare. Viceversa, si tratta di programmi utili per il cammino didattico dei più giovani, ed esclusivamente per quelle tipologie di testo fortemente omologabili come una lettera formale, o un resoconto tecnico.

I *software* di ausilio alla scrittura sono numerosi, e l'impressione che generano è che, per quanto variegati nell'interfaccia, nella ricchezza delle possibilità offerte, nel mercato al quale si indirizzano, essi tendano fondamentalmente a far compiere al computer il ruolo di un insegnante, che analizza, corregge, consiglia. Un insegnante terribile, però, perché rigido e cocciuto, assolutamente risoluto nell'imporre i modelli di scrittura predefiniti dal programma. Si cerca insomma di far fare al computer ciò che l'essere umano fa meglio, e questo contrasta con il principio generale già enunciato, rivelandosi inoltre, almeno dal punto di vista didattico, piuttosto inefficace.

Un approccio interessante perché diverso è quello suggerito da Hovstad, che infatti scrive "the project does not aim to replace teachers with computers but to help students"⁶⁴. Il computer in questo caso non ha pretese 'contenutistiche' (e questa umiltà dei programmatori è un bene prezioso), ma

⁶⁴ **Unni Hovstad**, "Computer Assisted Essay-writing: an Interdisciplinary Development project at Eikeli Grammar School", in Noel Williams, Patrick Holt, *Computers and Writing: models and tools*. Exeter, Intellect, 1989, p.150.

si limita a fornire indicazioni sul vocabolario, la grammatica, l'ortografia, la concordanza, il tempo, la punteggiatura. Il programma è in grado di controllare il testo, di evidenziare gli errori e di contarli per tipo, evidenziando i problemi maggiormente riscontrati. È ovvio che il programma 'fallisce' tutte le volte che si azzarda a fornire informazioni e consigli che non appartengono propriamente al piano espressivo, come quando giudica la complessità di una frase. Chi ha stabilito, infatti, che una frase lunga è necessariamente complessa? Non c'è un collegamento diretto, come il computer (o meglio, chi ha scritto il programma di scrittura) vorrebbe farci credere.

I due sistemi descritti finora operano sul testo già prodotto. Da questo punto di vista, non influenzano particolarmente il recupero dell'informazione, perché si limitano a correggere il testo, senza aggiungere informazione che potrebbe favorirne il recupero, e senza modificarne particolarmente struttura. Qualcosa di più fanno quei programmi, come il *Writer Assistant*⁶⁵, che accompagnano lo scrittore durante il processo di scrittura, anziché affiancarsi in fase di revisione quando il testo prodotto è ormai completo. I programmi di questo tipo sono definiti "out-line processors", in quanto assumono la struttura di un testo come una caratteristica permanente del testo, la visualizzano e permettono di modificarla⁶⁶. In questo modo lo scrittore può operare su vari livelli di un testo, scrivendone le varie parti e poi modificandone la posizione o la dimensione grazie ad una visuale ampia, o, come si dice in gergo, "a volo di uccello". Il *Writer Assistant* descritto dal saggio citato nasce dall'analisi dei processi mentali dello scrittore: si opera in particolare non sulle strutture mentali, inconoscibili, ma sulle rappresentazioni esplicitamente 'esterne' che lo scrittore crea e che quindi sono osservabili. Questa analisi conduce alla costruzione di una griglia, che si propone di mostrare lo spazio rappresentazionale della scrittura: non organizzata, organizzata non linearmente, organizzata linearmente. Si potrebbe discutere sull'efficacia di questo modello: qui ci limitiamo a dire che sulla base di questa modellizzazione cognitiva, efficace o grossolana che sia giudicata, si forniscono all'elaboratore una serie di modelli di scrittura, verso i quali 'orienterà' lo scrittore durante la produzione testuale. Il sistema sfrutta bene

⁶⁵Mike Sharples, James Goodlet, Lyn Pemberton: "Developing a Writer's Assistant", in Noel Williams, Patrick Holt, op.cit.

⁶⁶Cfr. Jay David Bolter, *Lo spazio dello scrivere. Computer, ipertesti e storia della scrittura*. Milano, Vita e Pensiero, 1993, p. 24.

le potenzialità del computer perché permette al lettore di osservare il suo testo così come appare mentre viene prodotto, ma fornisce anche una visione della struttura schematica dello stesso (fatta per esempio di un box chiamato 'premesse', collegato ai box di 'sfondo' e 'conclusioni', eccetera) e una terza visione, sempre strutturale ma non lineare, in cui il lettore vede il testo come una serie di 'isole' di contenuto più o meno connesse. L'utente può operare su tutti e tre i piani: modificare la struttura lineare, aggiungere nuove annotazioni e informazioni sulla struttura a rete, modificare direttamente il testo. Il computer provvederà ad operare una revisione periodica dell'opera, avvisando l'utente quando non sta rispettando i limiti strutturali che si è imposto (per esempio: l'introduzione è troppo lunga). Si può rilevare come questo sistema faciliti non poco la capacità di metacognizione dell'utente, che può visualizzare la struttura del suo testo, e rendersi consapevole dell'equilibrio fra le parti. Gli autori del saggio non ci dicono se questa struttura è 'recuperabile': se cioè un altro utente può accedere alla struttura lineare o reticolare prodotta dall'autore come base del testo. Se ciò fosse possibile, si fornirebbe uno strumento prezioso di recupero dell'informazione.

A conclusione di questa sezione, ribadiamo il più volte ripreso concetto che l'interazione tra utente e macchina funziona bene se ognuno riveste quei ruoli in cui riesce meglio: i sistemi esperti che favoriscono la scrittura 'funzionano' bene quando al computer è richiesto il recupero di occorrenze, la memorizzazione e l'applicazione di regole rigide e utilizzabili senza ambiguità; oppure, come nel terzo caso analizzato, quando al computer è richiesta la visualizzazione di un processo creativo che è gestito dall'utente, per quanto parta da un modello predefinito. L'interazione 'zoppica' quando si richiede al computer più di quanto sia in grado di fare: analisi dei contenuti, soluzione di problemi imprevisti. Non paia 'fuori luogo' questa dissertazione. Essa, come ogni dissertazione svolta nella prima parte di questa ricerca, ci tornerà utile nella seconda parte, quella dell'analisi di un caso specifico. Infatti, sulla scorta di quanto detto qui, si comprenderà meglio perché, per realizzare un testo che favorisca il successivo reperimento dell'informazione, alcuni dei processi produttivi sono stati affidati agli operatori umani (la ricerca dell'informazione pertinente), altri sono stati affidati al computer (la memorizzazione di grossi quantitativi di informazione). Procediamo ora allargando la nostra nozione di 'testo' alle nuove forme di testualità che l'elettronica ci offre.

2.2. Le caratteristiche multimediali del testo elettronico: l'ipermedia

Ricerca informazione testuale oggi non significa più soltanto sfogliare cataloghi, riviste, elenchi. Significa avere a che fare con forme di testualità molto più complesse e sofisticate di quelle consentite dalla scrittura lineare tradizionale. È perciò opportuno soffermarsi a considerare le caratteristiche di quell'universo testuale in continua espansione che viene solitamente definito ipermedia, per comprendere come il ricercatore di informazione debba porsi rispetto a questi strumenti. Il presente paragrafo si mostrerà più ricco di quelli che lo hanno preceduto: ciò è dovuto al fatto che dalla fine degli anni ottanta ad oggi il numero di ricerche dedicate all'ipertestualità è stato enorme, come Meyrowitz fa ironicamente notare, anche se non sempre a tanta quantità è corrisposta altrettanta qualità⁶⁷.

Il termine 'ipermedia' è la traduzione italiana dell'inglese '*hypermedia*', a sua volta derivato dall'unione dei termini *hypertext* e *multimedia*: a primo acchito questo concetto sembrerebbe comprendere tutte le forme di comunicazione strutturate in maniera ipertestuale in grado di veicolare contemporaneamente i contenuti di due o più media. Per approfondire l'analisi terminologica, cominciamo dal secondo termine, 'multimedia'. La multimedialità, letteralmente, indica la presenza di due o più media; più precisamente, la combinazione di due o più media continui, che debbono essere riprodotti in sincrono durante un ben definito intervallo di tempo. Per questo la televisione, che trasmette audio e video continui, è multimediale; un libro, che pure è costituito da testo e immagini, non lo è⁶⁸. Il concetto di ipertesto merita qualche maggiore riflessione.

La maggior parte degli studiosi è concorde nel far risalire l'idea di ipertesto a Vannevar Bush, autore del saggio "As we may think"⁶⁹, datato 1945. Da questo testo si evince come l'idea di ipertestualità sia nata con lo scopo di favorire la ricerca scientifica, la catalogazione delle informazioni, il loro recupero. Anziché conservare informazioni in archivi chiusi, compartimenti separati, ci si sforza di abbattere queste barriere (non solo fisiche, ma anche cognitive) e si permette la creazione di collegamenti trasversali. Nato, per così dire, come un sostegno all'apprendimento, al

⁶⁷ Norman Meyrowitz, "Hypertext- Does It Reduce The Cholesterol, Too?", in Nyce M. James, Kahn Paul (eds.), *From Memex to Hypertext. Vannevar Bush and the Mind's Machine*, San Diego, Academic Press, 1991.

⁶⁸ Andrew S. Tanenbaum, *Reti di computer*, Torino, Utet, 1997, p. 202.

⁶⁹ Vannevar Bush, 1945, *As we may think*, 1945 <http://www.isg.sfu.ca/~duchier/misc/vbush>

commento, alla memorizzazione e alla ricerca dell'informazione, l'ipertesto ne è diventato uno dei maggiori terreni di sfida, perché accrescendo le potenzialità creative degli autori, ha enormemente accresciuto, di converso, la necessità di 'porre ordine' da parte dei lettori. Anche Landow, rileggendo Bush e Nelson, sostiene che l'ipertesto sia nato perché la conservazione dell'informazione in un formato fisso e immutabile come quello lineare a stampa rende le operazioni di *information retrieval* piuttosto difficoltose.⁷⁰

Nel saggio di Bush si sottolinea la crescita dei materiali di ricerca prodotti dagli scienziati di tutto il mondo, e la difficoltà del ricercatore nel dover confrontare una mole enorme di risultati, statistiche, proposte. Il problema della gestione di eccezionali quantitativi d'informazione, insomma, ha origini più remote di quanto non fosse lecito sospettare. Ebbene, secondo Bush la risposta a questi problemi è una macchina di sua progettazione, il Memex (Memory Extender⁷¹). Si tratta di una scrivania che attraverso un ingegnoso sistema meccanico permette di proiettare testi, creare collegamenti, copiarne parti, attraverso la tecnologia che all'epoca pareva più promettente, quella dei microfilm. Questa idea di poter collegare documenti diversi, evidenziarne una parte, porre in secondo piano un'altra, è fondamentale per la ricerca scientifica, ed è l'idea alla base della nascita dei sistemi ipertestuali: si tenga presente, e su questo punto torneremo in seguito, che il progetto del *web* è l'opera di un fisico, non di uno scrittore. Bush, tuttavia, non parla ancora di ipertesto; il termine è di Ted Nelson, l'autore che per decenni si è occupato della progettazione di una rete di documenti che rendesse possibile collegare qualsiasi scritto della letteratura mondiale, il progetto Xanadu. Nelson aveva costruito il suo progetto intorno all'idea di ipertestualità e di rete: Xanadu avrebbe potuto permettere ciò che il *medium* della carta stampata, rigido e fortemente 'chiuso', aveva impedito alla letteratura di ogni tempo: manifestare apertamente i rimandi interni al testo le interconnessioni tra testi diversi.

Siamo alla ricerca di metodi che ci consentano di manipolare, sempre che riusciamo a scoprire come definirla, la 'vera' struttura di un testo. Ciò che noi forniremo è un servizio per immagazzinare

⁷⁰ **George P. Landow**, "The Convergence of Contemporary Critical Theory and Technology", in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.94.

⁷¹ Il termine è traducibile in italiano con "estensore di memoria" e rende bene l'idea di uno strumento che più che fornire all'utente nuove competenze cognitive, amplifica ed estende quelle già possedute.

e reperire porzioni arbitrarie di questa 'vera' struttura. La struttura di cui stiamo parlando è la letteratura.⁷²

Attualmente, purtroppo, Xanadu è ancora un progetto, e sebbene alcune sue caratteristiche siano state fatte proprie dal *web*, quest'idea di letteratura elettronica non si è ancora realizzata. Non si possono d'altronde nascondere alcune inquietanti prospettive che l'eventuale realizzazione di un sistema simile a Xanadu presenterebbe: una enorme rete testuale fatta di documenti collegati dagli utenti farebbe correre il rischio di accrescere in maniera esponenziale quella che Bloom⁷³ definisce l'ansia dell'influenza poetica, la consapevolezza cioè dello scrittore di essere un cattivo interprete (*a misinterpet*) della tradizione letteraria che l'ha preceduto. Sarebbe possibile evidenziare impietosamente queste influenze con un semplice *link*, e la paura di non essere capaci di scrivere nulla di realmente originale diventerebbe probabilmente un'ossessione.⁷⁴ D'altronde questo scenario è riproposto in ogni caso dal testo elettronico, come ricorda Bolter, un autore di cui ci occuperemo in seguito, quando scrive, riferendosi appunto alla teoria di Bloom per cui ogni autore travisa i suoi predecessori per creare un nuovo testo, essendo il fraintendimento una parte essenziale del processo di lettura:

Il computer dà espressione concreta all'atto della lettura (o del fraintendimento) intesa come interpretazione, e sfida il lettore a lottare con l'autore per il controllo dello spazio di scrittura.⁷⁵

Ci sia concesso di prenderci la libertà per una piccola digressione 'sociologica': nessuno ha mai acquistato un Memex, semplicemente perché se tecnologicamente era possibile costruirlo, economicamente si rivelava un'impresa disastrosa; e nessuno finora ha navigato in Xanadu⁷⁶, se non i progettisti e pochi sperimentatori. Bush e Nelson hanno svolto un ruolo di

⁷² **Ted Nelson**, *Literary Machines 90.1.*, Padova, Muzio Editore, 1992, p. 2/4 -2/6.

⁷³ **Harold Bloom**, *The Anxiety of Influence. A Theory of Poetry*, New York, Oxford U.P., 1973.

⁷⁴ Già attualmente il *web* è talmente ricco di documentazioni che si fa fatica a stabilire chi sia l'autore di un articolo, di una fotografia o di un'immagine computerizzata e chi siano coloro che l'hanno semplicemente 'copiato' sulle loro pagine. Figurarsi se oltre al plagio vero e proprio si considerassero anche le influenze inconse.

⁷⁵ **Jay David Bolter**, *Lo spazio dello scrivere. Computer, ipertesti e storia della scrittura.*, Milano, Vita e Pensiero, 1993, p.196.

⁷⁶ **Ted Nelson**, op.cit.

'*visionary*', cioè di precursori, senza veder realizzati i loro progetti perché mancava un contesto sociale pronto ad accoglierli ed ad attuarli. Contesto che invece ha celebrato il trionfo del *web*. Ma torniamo adesso ad alcune definizioni di ipertesto:

A term coined by Ted Nelson around 1965 for a collection of documents (or 'nodes') containing cross-references or 'links' which, with the aid of an interactive browser program, allow the reader to move easily from one document to another⁷⁷.

Con ipertesto intendo scrittura non sequenziale, testo che si dirama e consente al lettore di scegliere; qualcosa che si fruisce meglio davanti a uno schermo interattivo⁷⁸.

Per ipertesto intendo semplicemente la scrittura non sequenziale. La composizione di una rivista, con brani sequenziali, figure inserite e riquadri è quindi un ipertesto. E tale è la prima pagina di un quotidiano(...)⁷⁹

(L'ipertesto è) testo composto da blocchi di parole (o immagini) connesse elettronicamente secondo percorsi molteplici in una testualità aperta e perpetuamente incompiuta descritta dai termini **collegamento, nodo, rete, tela e percorso**.⁸⁰

"Hypertext" is non-sequentially *linked* pieces of text or other information. If the focus of such a system or document is on non-textual types of information, the term hypermedia is often used instead. In traditional printed documents, practically the only such *link* supported is the footnote, so hypertext is often referred to as

⁷⁷ FOLDOC è un *computing dictionary* disponibile in rete. Indirizzo <http://wombat.doc.ic.ac.uk>

⁷⁸ **Ted Nelson**, op.cit., p. 0/2.

⁷⁹ **Ted Nelson**, op.cit., p. 1/17.

⁸⁰ **George P. Landow**, *Ipertesto, il futuro della scrittura*, Bologna, Baskerville, 1993.93, p.5.

"the generalized footnote."⁸¹

What is hypertext? Well, I am tempted to refer to it as something more than purely a reference tool. Hypertext is both function and form. Simply meaning, it is a product of recent computer technology that is in and of itself a new form, but also serves the long since known function of referencing. It is the object of new a literary form called "Interactive Fiction." It has the potential of being the most powerful referencing tool in the history of humankind. Hypertext is the result of new database technologies brought about by the advancement of the computer.⁸²

The hypertext is the set of *linkages* that make available the right information, in the right format, at the right time- Just In Time Information.⁸³

Le definizioni potrebbero continuare, ma riteniamo opportuno fermarci qui per non dilungarci troppo: d'altronde, queste definizioni sono sufficienti ad evidenziare la magmaticità della materia in questione, ancora lungi dall'essere rigidamente strutturata. Il dibattito, accademico e non, è piuttosto vivo, come dimostrano queste parole di Meyrowitz:

One can glean, by reading the conference proceedings and tracking the press, that other people think that hypertext is fundamentally a tool for teaching writing. There are people who think that hypertext is a reference tool, a way to put encyclopedia or dictionaries online. Some are focusing on hypertext as a way to do argumentation, as a way to focus and connect the different parts of an argument.⁸⁴

⁸¹ **Jacob Nielsen**, *Hypertext 87*. [Http://www.sun.com/950523/columns/alertbox/ht87.html](http://www.sun.com/950523/columns/alertbox/ht87.html)

⁸² **Gary Potter**, *A Post-Modern Type of Thing*.

[Http://mbhs.bertraum.k12.ny.us/~tnellen/hypertext.html](http://mbhs.bertraum.k12.ny.us/~tnellen/hypertext.html).

⁸³ **Ray McAleese**, *Hypertext. Theory into practice*, Exeter, Intellect, 1999, p.XV.

⁸⁴ **Norman Meyrowitz**, "Hypertext- Does It Reduce The Cholesterol, Too?", in Nyce M. James, Kahn

Meyrowitz continua la sua lista aggiungendo che per altri l'ipertesto è prevalentemente uno strumento per, nell'ordine, il lavoro di gruppo, la *fiction* interattiva, la narrativa, la critica letteraria, la raccolta e la conservazione libraria, per non parlare di chi evidenzia l'utilità di tale sistema per i professionisti quali i medici e gli avvocati. La conclusione alla quale Meyrowitz giunge è laconica: "My hypothesis is that hypertext is all of that, but hypertext is more"⁸⁵. Ce n'è abbastanza per cercare di capirne di più, partendo da colui a cui è attribuita la paternità del termine ipertesto, Ted Nelson. La definizione di Nelson è piuttosto ambigua, perché da una parte il suo concetto di ipertesto sembra limitarsi alla forma del contenuto, senza delinearne la sostanza: ipertesto, nella definizione di Nelson, è anche un quotidiano, anche un'enciclopedia. Dall'altra, però, l'autore sottolinea che la presenza di uno schermo interattivo favorisce la fruizione ipertestuale. Ebbene, l'uso comune sembra aver dato maggiore credito a questa seconda accezione del termine, che limita il campo dell'ipertesto alla sua possibilità di fruizione elettronica. Nel linguaggio scientifico è ormai diffusa la considerazione di ipertesto come di un testo composto da blocchi di parole connesse elettronicamente secondo percorsi molteplici, fruibili dal lettore solamente attraverso un *software* ed un *hardware* adatto. Se alle parole si aggiungono anche blocchi di immagini, suoni, animazioni, allora l'ipertesto può dirsi multimediale. Si tratta di una definizione che non si discosta molto da quella data da Landow⁸⁶, di cui riprende l'idea di percorsi molteplici e quindi di ruolo più attivo del lettore, ma che sottolinea la presenza di un supporto elettronico. Si elimina così dal raggio d'azione tutta una gamma di testi che non possono essere considerati 'ipertesti multimediali':

- Tutti i testi stampati su supporto cartaceo, quale che sia la loro struttura narrativa e indipendentemente dal ruolo più o meno attivo in essi realizzato dal lettore.

- Tutti i testi pensati originariamente come 'tradizionali' e in seguito riportati su supporto elettronico. Si può infatti creare un ipertesto multimediale sulla Divina Commedia, con tanto di fotografie e musiche, così come si può ricopiare l'opera con un *word processor* e digitalizzarla; nel

Paul (eds.), *From Memex to Hypertext. Vannevar Bush and the Mind's Machine*, San Diego, Academic Press, 1991, p. 287.

⁸⁵ Norman Meyrowitz, op.cit., p.288.

⁸⁶ cfr. George P. Landow, *Ipertesto, il futuro della scrittura*, Bologna, Baskerville, 1993.

Si noterà che abbiamo volutamente saltato una parte importante, quella che definisce l'ipertesto una testualità "perennemente incompiuta" (p.5). Questo può essere vero per determinati ipertesti *on-line*, non per quelli *off-line*.

primo caso abbiamo una nuova opera, che è altro dal testo dantesco, nel secondo abbiamo un testo elettronico non ipertestuale. Connessa a tale definizione di ipertesto è quella di collegamento o *link*, che attraverso il *cliccaggio*⁸⁷ di una *hotword*,⁸⁸ ci permette il passaggio immediato da una parte di un documento ad un'altra (*anchor*), da un documento ad un altro (*hyperlink*) e da una parte specifica di un documento ad una parte specifica di un altro documento (*hyperanchor*). La seguente tabella sintetizza rapidamente un confronto tra le caratteristiche dell'ipertesto e del testo tradizionale:⁸⁹

Testo tradizionale	Ipertesto
Sequenzialità	Non sequenzialità
Linearità	Multilinearità
Chiusura	Apertura
Passività	Attività- Interattività
Stabilità	Instabilità- Fluidità
Maggior numero di nessi impliciti	Maggiore esplicitazione dei nessi
Presenza di centro	Decentramento
Principio di organizzazione	Assenza di organizzazione
Staticità	Dinamicità
Fruizione individuale	Fruizione collettiva- Cooperazione

Tabella 2: principali differenze tra testo tradizionale ed ipertesto

Sarebbe possibile soffermarsi su ognuno di questi aspetti, e suscitare dei dubbi su quasi tutti i passaggi di questo schema. È proprio vero che il lettore tradizionale ricopra un ruolo passivo? Le ricerche portate avanti da molti studiosi post-strutturalisti sembrano andare in direzione diversa. Fino a che punto un ipertesto può essere definito non sequenziale, caratterizzato cioè da blocchi indipendenti il cui ordine non è importante ai fini della lettura? E poi,

⁸⁷ Non è una bella espressione ma non abbiamo trovato equivalenti adatti in italiano.

⁸⁸ Una *hotword* è una parola cliccabile identificabile dal colore diverso e dal cambiamento di forma che assume il puntatore del *mouse* quando le passa sopra. Alcuni testi stanno cominciando a tradurre questo termine con l'equivalente italiano 'parola calda', ancora poco diffusa.

⁸⁹ La tabella è stata riprodotta sulla base degli appunti del seminario sull'ipertestualità tenuto dalla dottoressa Giovanna Cosenza nell'ambito del corso di Semiotica del Testo tenuto all'Università di Bologna nell'anno accademico 1996-1997.

all'interno di questi blocchi, non vigono forse le tradizionali leggi della sequenzialità? Davvero un ipertesto è così decentrato? E allora perché quasi tutti i prodotti editoriali su CD-ROM presentano una pagina che svolge la funzione di indice, così come i siti web presentano sovente una struttura molto organizzata?

Viene da pensare anche a quello che scriveva Barthes anni fa⁹⁰, quando preannunciava il passaggio dall'opera al testo. La prima è un oggetto di consumo definito, singolare, con tanto di autore che evidenzia la distanza tra scrittura e lettura, moderatamente simbolico. Il secondo non è definito materialmente, è radicalmente simbolico, plurale (cioè dotato di una pluralità di significati), riduce le distanze tra lettura e scrittura, non è protetto dalla 'garanzia' dell'autore, semplicemente 'ospitato' dal testo. Barthes, metaforicamente, definisce l'opera un 'organismo', il testo 'una rete'. L'ipertesto tende a realizzarsi come 'testo', ma contiene dentro di sé tutte le tracce dell' 'opera'.

Un altro autore che ha fornito un importante contributo allo studio della testualità elettronica è senz'altro Jay David Bolter⁹¹. Anche Bolter fa risalire a Bush l'idea di ipertesto inteso come estensione delle capacità della mente, ma contestualizza il discorso in una più ampia disamina storica dello "spazio della scrittura": la nuova tecnologia si apparenta più strettamente al discorso orale che alla stampa convenzionale. Come il poeta orale si muoveva assemblando e collegando blocchi di formule imparati a memoria, così la scrittura elettronica si manifesta sostanzialmente come processo associativo. La scrittura elettronica è una scrittura "aforismatica", che produce blocchi di testo indipendenti, laddove la linearità dell'antica tradizione del libro a stampa richiedeva l'utilizzo di coordinate e subordinate. Questo è un concetto molto importante per il reperimento dell'informazione: il testo elettronico, creato in blocchi indipendenti di testo, può facilitare il recupero di un dato, semplicemente richiamando quel blocco inequivocabilmente individuabile. Bolter ricorda come molti autori di questo secolo abbiano avvertito i limiti soffocanti del testo lineare a stampa e abbiano cercato di vincerne le costrizioni pur agendo all'interno del sistema lineare. Un esempio perfetto di ipertesto *ante litteram*, in quest'ottica, sono le famose

⁹⁰ **Roland Barthes**, "From work to text", in *Textual Strategies*, edited by J.H. Harari, Ithaca, Cornell U. P., 1979.

⁹¹ **Jay David Bolter**, *Lo spazio dello scrivere. Computer, ipertesti e storia della scrittura*. Milano, Vita e Pensiero, 1993.

Ricerche Filosofiche di Wittgenstein⁹², con la tipica struttura di annotazioni collegate tra loro poco rigidamente. La scrittura elettronica è inclusiva, “è un *continuum* nel quale possono felicemente coesistere sistemi plurimi di rappresentazione”⁹³, al contrario della scrittura alfabetica, che per quanto possa essere contaminata dalla presenza di simboli ‘estranei’ come i numeri arabi, “continua a mirare ad una semplificazione per esclusione”. Una scrittura davvero multimediale, che faccia uso delle potenzialità offerte dall’utilizzo di supporti elettronici, può quindi ‘contaminare’ il testo con più facilità di quanto non possa la scrittura lineare tradizionale. Questa ‘contaminazione’ del testo con elementi estranei appartenenti ad altri codici, in particolare quello visivo, ne permette una migliore gestione e soprattutto un più facile recupero, come vedremo quando parleremo dei sistemi di *mark-up*. Lo studio di Bolter è molto affascinante perché mette in luce i legami cognitivi che la nuova tecnologia ha con tecnologie più antiche, dimostrando che la scrittura elettronica non è un elemento di discontinuità nella storia della scrittura ma una logica evoluzione. La ‘scrittura per immagini’, per esempio, è una forma di scrittura primordiale: si faccia riferimento alle incisioni risalenti ad epoche preistoriche che rappresentano scene di battaglia o avvenimenti con un determinato ordine cronologico. Al contrario di quella alfabetica che è secondaria perché rimanda al sistema fonetico, questo tipo di scrittura è primaria, pertanto indipendente dal testo parlato. La scrittura per immagini è meno precisa di quella alfabetica, perché è più vicina al lettore, al quale sono concesse un’ampia gamma di interpretazioni, che allo scrittore. Ebbene, a questo tipo di scrittura ‘primordiale’ sembra ritornare la scrittura elettronica, che combina il linguaggio per immagini (spesso metaforico, iconico, come si è detto in precedenza) con il linguaggio alfabetico, riuscendo a farli convivere. Studiando bene il nostro passato, sembra suggerirci l’autore, comprenderemo meglio il nostro presente. Scrive ancora Bolter:

Il computer libera l’autore dall’artificio consunto della scrittura lineare, ma ad un prezzo: egli deve permettere l’intervento del lettore nello spazio della scrittura.⁹⁴

Questa frase dovrebbe diventare uno dei ‘comandamenti’ dello scrittore

⁹² Ludwig Wittgenstein, *Ricerche filosofiche*, Torino, Einaudi, 1967.

⁹³ Jay David Bolter, op.cit., p.77.

⁹⁴ Jay David Bolter, *Lo spazio dello scrivere. Computer, ipertesti e storia della scrittura*. Milano, Vita e Pensiero, 1993, p.186.

multimediale; soprattutto essa conferma la tesi per cui lo scrittore ‘elettronico’ non può costringere (a meno che non si tratti di una scelta consapevole: ma questo è un altro discorso) il lettore ad un unico percorso di lettura, e deve strutturare il suo elaborato in maniera da facilitare una lettura a balzi, una facile consultazione del testo ed un recupero immediato delle informazioni indispensabili. Sono stati questi gli obiettivi che, come curatori del testo che analizzeremo nella seconda parte, ci siamo posti, con lo scopo di ‘invitare’ l’intervento del lettore in uno spazio in cui è chiamato ad essere protagonista e non semplice spettatore. Se il testo cartaceo è una “macchina pigra” che ha bisogno di un lettore attivo per funzionare⁹⁵, il testo elettronico è una “macchina inerte”, dal momento che non si muove senza l’intervento deciso del lettore, cui deve essere lasciato il giusto spazio.

Per comprendere appieno l’ampia rosa di discorsi esistenti sull’ipertestualità bisogna considerare la logica commerciale che accompagna sin dalla sua nascita la tecnologia ipertestuale, logica che ha ostacolato un approccio oggettivo all’introduzione di questa forma di testualità nel panorama mediatico. Così le comparazioni tra ipertesto elettronico e testo tradizionale a stampa hanno spesso finito per magnificare gli aspetti innovativi del primo, senza riflettere sul fatto che ipertestualità e intertestualità (quest’ultima in relazione alla natura ipertestuale che la Rete delle Reti ha assunto in seguito alla nascita del *web*) sono innanzitutto modalità di fruizione di un testo, e solo in seconda battuta ‘tecnologie’ testuali. Anche la testualità tradizionale si presta a essere ‘processata ipertestualmente’, con un lettore che opera spostamenti a livello metatestuale, fuggendo la rigidità, a questo punto pregiudizialmente ipotizzata tale, della struttura del testo a stampa. Analogamente, la presunta apertura degli ipertesti elettronici, la posizione di predominio delegata al lettore/utente nell’attività di ‘produzione del testo’, vengono notevolmente ridimensionate se si analizzano con attenzione i vincoli imposti dall’autore nella strutturazione del testo stesso; la libertà del lettore diviene sotto questa luce niente più che una serie di scelte condizionate da una precisa strategia. Tale concetto è espresso chiaramente da Pellizzi, quando scrive:

(...)un ipertesto, in quanto organizzazione logica di connessioni, è, diversamente da quanto sembra dalle promesse di libertà incondizionata, un’entità dittatoriale. Ove manchino gli strumenti

⁹⁵ La definizione è di Eco. Cfr. **Umberto Eco**, *Lector in fabula*, Milano, Bompiani, 1979, p.24.

logici di orientamento cui si è fatto cenno, tale aspetto dittatoriale non viene affatto meno: ne aumenta soltanto l'occultamento.⁹⁶

Ridefinire lo *status* del lettore in termini di maggiore o minore libertà di interazione nell'ipertesto rispetto al testo scritto tradizionale conduce, inevitabilmente, ad un *impasse*. Il ruolo del lettore si mostra, infatti, in bilico tra un 'navigatore' in balia dell'ipertesto, che ha solo la possibilità di perdersi per rotte già tracciate tra le varie isole di testo, e un 'costruttore di testo', che stabilisce i nessi di causalità tra le pagine, e che si fa quasi co-autore. Una maggior ricchezza dell'interfaccia (e del supporto fisico in generale) non corrisponde necessariamente ad una maggiore potenzialità di interazione, così come la libertà di un lettore non si misura in termini di numero di percorsi di lettura già presenti nel testo: tante strade 'preconfezionate' non fanno un lettore più 'potente'. A riguardo Cadioli si mostra alquanto critico nei confronti degli "entusiasmi poststrutturalisti" che hanno ritrovato nell'ipertesto gli estremi per decretare la morte definitiva dell'autore e la libertà assoluta del lettore, ricordando che l'interattività ipertestuale si sviluppa proprio nell'equilibrio tra la presenza non secondaria dell'autore e la libertà di scelta offerta al lettore.⁹⁷ Non è opportuno qui proseguire oltre questo dibattito: per quanto concerne l'attività di ricerca dell'informazione rilevante, con l'ipertesto essa è facilitata solo nei termini in cui l'autore ipertestuale si è sforzato di facilitarla. Un ipertesto, infatti, fa uso di tecnologie che permettono una ricerca agevolata dell'informazione, ma solo attraverso percorsi già predisposti dall'autore; per tutti gli altri percorsi 'alternativi', il recupero è anche più complesso che per il testo tradizionale. Per cui, il produttore di un testo che voglia facilitarne la fruizione da parte del lettore e il recupero informativo, dovrà riflettere a lungo prima di scegliere la modalità di scrittura ipertestuale, e in ogni caso non potrà essere accusato di arretratezza se preferirà affidarsi alle modalità testuali lineari.

Prima di procedere è opportuno soffermarsi sulla definizione di ipermedia, di cui abbiamo già visto un accenno. Nelson, probabile coniatore anche di questo termine, non ne dà una definizione precisa, ma lascia intendere che tra gli ipermedia vanno considerati videodischi interattivi, CD-ROM, CDI, DVI.

⁹⁶ **Federico Pellizzi**, "Letteratura on-line. Come si fa un bollettino elettronico", in Patrizia Bellmann Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.330.

⁹⁷ **Alberto Cadioli**, *Il critico navigante [saggio sull'ipertesto e la critica letteraria]*, Genova, Marietti, 1998, p. 93.

In un certo senso il termine ipermedia, che l'autore usa soltanto al plurale, non è più una semplice concezione astratta e quindi applicabile a vari media come l'ipertesto, ma una definizione comprendente anche i supporti materiali.

Il termine ipermedia estende semplicemente l'idea di testo dell'ipertesto includendovi informazioni visive, suono, animazione e altre forme di dati. (...) Nelle pagine che seguono, userò i termini ipertesto e ipermedia in maniera intercambiabile.⁹⁸

È utile considerare 'Ipermedia' come un ibrido formato da due metà ben rappresentate del nome stesso: la prima metà 'iper-' proviene da 'ipertesto', un'idea teorica non particolarmente nuova che però solo recenti sviluppi tecnologici hanno permesso di implementare; la seconda metà, '-media', proviene invece da 'multimedia', che, al contrario della prima, non è un'idea teorica, ma una tecnologia molto recente basata sullo sviluppo dei calcolatori. È facile, come fanno molti, farsi trascinare da questa combinazione terminologica e 'sostanziarla' così da considerare ipermedia la somma di ipertesto e multimedia (quando addirittura non si consideri, come spesso accade anche in pubblicazioni che aspirano ad essere scientifiche, di usare ipermedia semplicemente come sinonimo di multimedia o come semplice variante di esso); di somma invece non si tratta, Ipermedia non è semplicemente un Ipertesto cui sono stati aggiunti altri media, né un multimedia che presenta legami ipertestuali. È qualcosa di più e insieme di specifico rispetto a queste cose: l'ibrido è una nuova specie, in cui le parti componenti non si sommano ma, restando nella metafora, si moltiplicano. Ipermedia è l'integrazione dei media in un unico, nuovo oggetto comunicativo non riferibile a, né comprensibile in, nessuno dei singoli media

⁹⁸ **George P.Landow**, *Ipertesto, il futuro della scrittura*, Bologna, Baskerville, 1993, p.6.

specifici componenti.⁹⁹

Anche il termine ipermedia presenta alcune sfumature non ben delineate, perché per Nelson e Landow, come abbiamo visto, a volte diviene sinonimo di ipertesto, a volte identifica l'*hardware* che ne permette la lettura. La definizione di Antinucci (condivisa da molti, tra l'altro) sottolinea l'integrazione di molti media in uno soltanto. Ma come giustamente fa notare Blasi¹⁰⁰, questa definizione non è specifica, perché anche la televisione, ad esempio, presenta elementi di sincretismo, cioè compresenza di molteplici tipologie testuali. Sulle riviste specializzate e non, spesso si parla di ipermedia in riferimento a reti digitali come Internet, a sistemi di trasmissione dati analogici o digitali come il satellite, a ipertesti multimediali, a nuovi servizi di comunicazione disponibili via cavo (*video on demand, pay-tv, pay per view*, ecc.), a nuovi media che cercano di conquistare il mercato. Siccome si tratta di un insieme avente estensione vasta e disomogenea, risulta difficile delineare con chiarezza l'intensione di 'ipermedia'. Tale mancanza di precisione indica bene il fermento e l'elasticità che caratterizzano questi settori, che devono ancora sedimentare un corpus di conoscenze acquisite.

Dal punto di vista della nostra ricerca, quello che è importante sottolineare è che la confluenza di vari codici in un solo strumento di comunicazione, che si chiami ipertesto multimediale o ipermedia poco interessa, arricchisce enormemente le possibilità di lettura, ma al tempo stesso richiede una maggiore consapevolezza da parte dell'autore del testo dei rischi che il suo lavoro comporta. Maggiormente ricca, variegata e complessa è l'informazione, maggiore è il bisogno di efficaci strumenti di *information retrieval* che limitino lo spaesamento del fruitore. Il primo a mettere in guardia di fronte ai rischi che corre la produzione ipertestuale è stato Conklin, nel 1987, esprimendo una giusta preoccupazione nei confronti del disorientamento e del sovraccarico cognitivo del lettore. Il disorientamento consiste nella perdita di cognizione della posizione in cui ci si trova all'interno di un testo elettronico e di come la si è raggiunta, causato dall'ampia gamma di direzioni in cui ci si può muovere e dal relativamente

⁹⁹ **Francesco Antinucci**, *Ipermedia e processi di apprendimento*
[Http://www.infosys.it/INFO90/obbligo/antinucc.html](http://www.infosys.it/INFO90/obbligo/antinucc.html)

¹⁰⁰ **Giulio Blasi**, *Dentro e fuori di Internet, osservazioni sulla nascita di un nuovo medium. Dispense per il corso di Teoria e tecnica degli ipertesti multimediali*, Università di Bologna, A.A. 1997-98.

alto grado di libertà che è concesso al lettore.

Getting "lost in space." Along with the power to organize information much more complexly comes the problem of having to know (1) where you are in the network and (2) how to get to some other place that you know (or think) exists in the network. I call this the disorientation problem. Of course, one also has a disorientation problem in traditional linear text documents, but in a linear text, the reader has only two options: he can search for the desired text earlier in the text or later in the text. Hypertext offers more degrees of freedom, more dimensions in which one can move, and hence a greater potential for the user to become lost or disoriented. In a network of 1000 nodes, information can easily become hard to find or even forgotten altogether.

Il sovraccarico cognitivo è invece un fenomeno causato dalla pressione sul lettore generata dalla necessità di dover compiere continuamente delle scelte che, viceversa, in un testo lineare sono effettuate dall'autore. In particolare, il lettore è 'miope', perché non può prevedere l'importanza di un eventuale collegamento se non dopo averlo realizzato. Ogni qualvolta si ferma per decidere se spostarsi o meno sul collegamento suggerito, egli viene inevitabilmente distratto dal testo che stava leggendo.

The problem of cognitive overhead also occurs in the process of reading hypertext, which tends to present the reader with a large number of choices about which links to follow and which to leave alone. These choices engender a certain overhead of metalevel decision making, an overhead that is absent when the author has already made many of these choices for you. At the moment that you encounter a link, how do you decide if following the side path is worth the distraction? Does the label appearing in the link tell you enough to decide? This dilemma could be called "informational myopia."

Ad oltre dieci anni da questo articolo, con una notevole evoluzione

tecnologica degli ipertesti alle spalle, le questioni acutamente sollevate da Conklin sono tutt'altro che di secondo piano. Conklin intravede tre tipi di soluzioni che possono ridimensionare, anche se non risolvere del tutto, il problema: velocizzare i tempi di visualizzazione di un collegamento; fornire una breve spiegazione, visualizzata in una finestra alternativa al testo che non si sovrapponga su quest'ultimo, che giustifichi la possibilità di collegarsi ad un'altra sezione; organizzare un *browser* grafico che mostri schematicamente la porzione di struttura ipertestuale che si sta percorrendo.

Ci sono inoltre altre proposte per la risoluzione di questo genere di problemi. Innanzitutto, l'uso del linguaggio metaforico per costruire le interfacce interattive. Le metafore utilizzate privilegiano soprattutto lo spazio visivo: così l'ipertesto diventa uno spazio da esplorare (*browser*) senza una strategia di ricerca precisa, oppure uno spazio in cui viaggiare (*travel*)¹⁰¹: il termine presuppone in più la varietà di mezzi di trasporto utilizzabili. Lo strumento metaforico visivo per eccellenza diventa la mappa grafica, benché, come ricorda McAleese:

Of course, with only a few nodes, it is simple to show them in one map or diagram. With thousands of nodes new solutions are called for.¹⁰²

In casi più complessi, è consigliabile fornire all'utente più strumenti di rappresentazione cognitiva, che lo aiutino a focalizzare meglio la sua ricerca. Scrivono a riguardo Edwards e Hardmann¹⁰³ che una scelta consigliabile è quella di offrire più modalità di accesso all'informazione: quella ipertestuale, che favorisce l'utente meno esperto attraverso i percorsi predefiniti dall'autore; quella gerarchica, che fornisce una struttura di argomenti suddivisi per categorie ed indici; quelle miste, che coinvolgono aspetti gerarchici e aspetti ipertestuali. L'utente in questo modo può prendere coscienza dell'ambiente testuale in cui si muove attraverso il sistema di ricerca che preferisce, e decidere di passare ad un sistema diverso solo quando ha ormai visionato, seppure per sommi capi, i confini del testo in esame.

¹⁰¹ **Lesley Allinson, Nick Hammond**, *A learning support environment: the hitch-hiker's guide*, in McAleese Ray, *Hypertext. Theory into practice*. Exeter, Intellect, 1999.

¹⁰² **Ray McAleese**, op.cit., p. IX.

¹⁰³ **Deborah M. Edwards, Lynda Hardman**, *'Lost in Hyperspace': cognitive mapping and navigation in a hypertext environment*, in Ray McAleese, op.cit.

Ma le difficoltà che l'utente di un testo elettronico incontra non si fermano qui. Cadioli¹⁰⁴, autore che abbiamo già citato e che si occupa del problema dal punto di vista del critico letterario, lamenta soprattutto una diffusa genericità nel testo elettronico che un discorso filologico non può tollerare. La scomparsa degli elementi che Genette definirebbe peritestuali¹⁰⁵ nell'editoria elettronica porta ad una amalgama indefinito di testualità in cui si trascura l'importanza di quegli elementi tipografici che rappresentano degli "strumenti di bordo" necessari per la navigazione. Un critico letterario, ma anche un qualsiasi ricercatore accurato, non può trascurare l'assenza da molti documenti elettronici di informazioni riguardanti l'anno di realizzazione, l'edizione, la versione presentata, la collocazione di una collana, le indicazioni dei curatori, le prefazioni, la composizione grafica (e quindi i formati di carattere usati, il numero di righe per pagina, per esempio): il testo elettronico non ha insomma alcuno "stato", cioè forma tipografica. Se consideriamo il paratesto come l'espressione visibile delle scelte dell'autore e dell'editore, attraverso i quali quest'ultimi definiscono i contatti con la comunità di lettori che vogliono raggiungere, ci possiamo rendere conto dell'importanza, troppo spesso trascurata, che il paratesto ricopre anche per il testo elettronico. Lo stesso Cadioli, in un altro saggio¹⁰⁶, ribadisce la necessità della consapevolezza teorica dell'uso dell'ipertesto: forzando il concetto di paratesto, infatti, si può arrivare a concludere che in un ipertesto ogni blocco testuale svolga la funzione di paratesto per gli altri; e considerando poi Internet come un enorme ipertesto, ogni testo presentato in rete diventa paratesto di qualcos'altro. Il paradosso è presentato da Cadioli per mostrare la necessità di creare nuove categorie concettuali per i testi elettronici, senza forzare quelle già esistenti, che altrimenti corrono il rischio di essere svuotate di ogni significato. Lo studio dei testi elettronici denuncia insomma delle ampie 'falle' nella struttura complessiva, dei vuoti concettuali che non possono essere colmati con una semplice operazione di 'travaso' di

¹⁰⁴ **Alberto Cadioli**, *Il critico navigante [saggio sull'ipertesto e la critica letteraria]*, Genova, Marietti, 1998.

¹⁰⁵ **Gérard Genette**, *Soglie. I dintorni del testo*, Milano, Einaudi, 1989. Genette definisce il paratesto come una soglia, appunto, attraversata la quale si arriva al testo vero e proprio. Sono elementi paratestuali peritestuali i titoli, gli indici, le epigrafi, la prefazione, l'introduzione, la quarta di copertina, il *prière d'insérer*, la postfazione, la copertina, il frontespizio, le note e quant'altro si pone 'vicino' al testo, intorno ad esso, senza diventarne parte integrante. Elementi paratestuali epitestuali sono invece quei messaggi che si trovano all'esterno del libro, più isolati ma comunque collegati ad esso, di natura pubblica, come le presentazioni convegni, le interviste, le promozioni pubblicitarie, o di natura privata, come i diari personali degli autori, le corrispondenze, gli appunti.

¹⁰⁶ **Alberto Cadioli**, "Le soglie dell'ipertesto", in Bellmann Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.50.

concetti appartenenti ad altre sfere culturali, ma richiedono la scrittura di nuove regole *ad hoc*. Il testo elettronico, scrivono Bérard e Mordenti, viene proiettato su uno schermo mobile e impalpabile, lasciando molto spazio all'improvvisazione di autore e produttore, senza "quell'insieme di *regole* che, sedimentandosi per secoli nella nostra tradizione, conferiscono rigore e riconoscibilità alle edizioni cartacee"¹⁰⁷. Dal punto di vista della ricerca dell'informazione, la perdita di queste caratteristiche tipografiche e fisiche, disorienta non poco il ricercatore, perché rende difficile distinguere a colpo d'occhio in rete il saggio importantissimo di un premio Nobel dalla *homepage* personale di uno studente liceale, mentre in biblioteca non avremmo difficoltà a distinguere un libro brossurato da un tascabile economico. D'altra parte, questo è il risvolto della medaglia della presunta democraticità della rete, e non è del tutto disprezzabile: a volte un tascabile da mille lire contiene informazioni più preziose di un'edizione con copertina in pelle e caratteri d'oro.

A questo punto ci sentiamo in grado di esprimere la nostra opinione a riguardo dell'utilità delle tecniche ipermediali per la ricerca dell'informazione. A nostro parere esse si dimostrano efficaci quando lo scopo principale è fornire un'idea globale di un argomento, preferendo la generalità alla specificità; quando permettono l'esplicitazione di informazione altrimenti tacita, informazione che facilita la focalizzazione da parte del lettore delle sezioni più interessanti dei documenti. Questa esplicitazione può essere data da una mappa, meglio se tridimensionale, del documento; da un indice contenutistico; da una serie di annotazioni dell'autore stesso, i 'segnali' testuali di cui si è già detto. Con questi strumenti si rende visibile al lettore una struttura testuale altrimenti invisibile, se ne permette con più consapevolezza la gestione, e di conseguenza si rende possibile il passaggio immediato da una parte di contenuto ad un altro. In altri casi, il tradizionale testo cartaceo non è da disprezzare. Per lo studio e l'apprendimento, che spesso richiedono letture ripetute dello stesso testo, la carta è ancora un medium valido, perché affatica meno la vista. Allo stesso modo è preferibile lo strumento cartaceo nel caso di strumenti di consultazione di uso facile e immediato: è molto più comodo, in caso di dubbio, sfogliare un dizionario tascabile di sinonimi e contrari, che accendere il computer e avviare il programma adatto. Per non parlare del fatto

¹⁰⁷ Claude Cazalé Bérard, Raul Mordenti, "La costituzione del testo e la «comunità degli interpreti»", in Patrizia Bellman Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.15.

che il dizionario si trasporta ovunque praticamente senza i costi che comportano invece i calcolatori elettronici. Sul piano strettamente linguistico, la tecnologia attuale permette di integrare strumenti e tecniche diversi in un solo supporto fisico, di natura elettronica, il che è stato a lungo lodato come una svolta epocale nella storia del pensiero. Se però ci si sofferma a riflettere sulla questione senza farsi influenzare più di tanto dagli slogan di qualche produttore di *hardware* o *software*, si riconosce che quello che è sicuramente un vantaggio, economico ma anche comunicativo, non deve altresì diventare un obbligo. Nessuno impedisce di affiancare alla rapidità del recupero elettronico, il più tradizionale apprendimento su testo cartaceo; laddove il primo disorienta, il secondo può rassicurare; di converso, laddove il secondo rallenta, il primo velocizza. Se lo scrittore ha a disposizione vari media di comunicazione, perché deve limitarsi ad usarne solo uno? Meglio scegliere di volta in volta la tecnologia da attuare, a seconda delle esigenze comunicative e delle restrizioni del mezzo fisico. Scrive a proposito Meyrowitz che ciò di cui abbiamo realmente bisogno è un “ambiente informativo” (*information environment*) di cui l’ipermedia è un elemento importante e fondamentale, ma non l’unico. Come una dieta equilibrata conduce al benessere se fa uso di diversi alimenti, così questo ambiente informativo può favorire l’universo dell’informazione a patto di essere vario e bilanciato¹⁰⁸.

È bene allora affiancare le tecnologie diverse, facendole interagire, utilizzando ciascuna per i compiti per i quali è più adatta. Rimandiamo ancora una volta alla seconda parte della nostra indagine per un approfondimento della questione. A questo punto, visto che si è parlato a più riprese di tecnologie, è giusto soffermarci su queste ultime, sulle loro risorse, sui loro limiti.

¹⁰⁸ Norman Meyrowitz, “Hypertext- Does It Reduce The Cholesterol, Too?”, in Nyce M. James, Kahn Paul (eds.), *From Memex to hypertext. Vannevar Bush and the Mind’s Machine*, San Diego, Academic Press, 1991, p.317.

3. Alcuni strumenti informatici per l'*information retrieval*

"Let the data shape the database and not viceversa"

W.Horton

3.1. Editoria on-line ed editoria off-line

Prima di intraprendere qualsiasi analisi dei sistemi informatici di reperimento dell'informazione, è doveroso chiarire, a mo' di premessa, una classificazione riguardante la natura dell'editoria elettronica, ipertestuale e non. I testi elettronici possono essere suddivisi in due grandi gruppi, considerando come criterio selettivo la presenza o meno degli stessi in una rete:¹⁰⁹ i primi si dicono *on-line*, i secondi *off-line*. Da una parte ci sono allora tutti i siti sul *web*, dai giornali multimediali alle pagine gestite da enti pubblici, dalle *soap-opera on-line* alle pagine *html* create da singoli utenti. Dall'altra ci sono gli ipertesti la cui esistenza è più nitidamente caratterizzata da un supporto fisico, sia esso un CD-ROM, un *floppy* o un *hard-disk*. L'enorme differenza tra queste due famiglie di ipertesti consiste nel fatto che per i primi si può effettivamente parlare, almeno in alcuni casi, di testi 'aperti', la cui connessione con il resto della rete attraverso collegamenti ipertestuali (*link*) ne fa oggetti dai contorni meno definiti. Per i secondi questo discorso non vale: le letture possibili di un'enciclopedia multimediale sono numerosissime, potenzialmente infinite, ma comunque non si esce dai territori delimitati e rigidi del supporto. Se vogliamo, ci troviamo di fronte a quelle somiglianze di famiglia di wittgensteiniana memoria: il gruppo degli ipertesti *off-line* è l'anello che congiunge i testi tradizionali, di cui mantiene la rigidità dei contorni e la finitezza dei contenuti (prima o poi un'enciclopedia multimediale su CD-ROM l'avrò letta tutta), con gli ipertesti su rete, che invece potenzialmente possono rimandare direttamente ad altri ipertesti collegati. Ultimamente stanno nascendo degli 'ibridi', per esempio

¹⁰⁹ Landow trova ben quattro definizioni diverse di rete. (George P.Landow, *Iper testo, il futuro della scrittura*, Bologna, Baskerville, 1993, p.30-31). Quella utilizzata qui corrisponde per sommi capi alla terza delle sue definizioni, per cui una rete è "un sistema elettronico che comprende diversi computer, con relative connessioni via cavo che consentono alle singole macchine, alle stazioni di lavoro e alle unità di lettura e scrittura di condividere informazioni".

enciclopedie che si aggiornano collegandosi a Internet, o siti che sono più chiusi e rigidamente delimitati di un giornale mercantile del XVIII secolo. Un'altra differenza importante, come ricorda Cadioli¹¹⁰, consiste nel fatto che la ricerca dell'informazione *off-line* è facilitata dalla presenza di supporti fisici che in qualche modo costituiscono il paratesto di questi prodotti, paratestualità assolutamente assente in rete: un CD-ROM, o un *floppy disk*, ha infatti una copertina e un retrocopertina che forniscono informazioni sul suo contenuto; può essere catalogato in biblioteca e archiviato proprio come se fosse un libro.

Le potenzialità ipermediali realizzabili su supporti fisici (in particolare su CD-ROM) *off-line*, sono di gran lunga superiori rispetto a quelli in rete: ciò è dovuto all'indiscutibile dato di fatto che la rete attuale è intrinsecamente lenta, piena di 'strozzature' che ostacolano il flusso dell'informazione e, ad oggi, impediscono una rapida consultazione delle pure enormi risorse disponibili; questo discorso è valido in maniera particolare per il materiale audiovisivo, ossia per quello 'più pesante' da trasportare.

Da un punto di vista operativo è abbastanza semplice capire quali siano le differenze tra l'editoria elettronica *on-line* e quella *off-line*. Infatti, alla ricca veste grafica (talvolta al limite dell'eccesso) dei prodotti *off-line* si contrappone la forzata essenzialità delle risorse *web*. Non dobbiamo dimenticare, infatti, che tra la qualità dell'interfaccia grafica *on-line* e quella dei più comuni sistemi operativi esiste un ritardo quantificabile in alcuni anni di evoluzione del *software*, che solo in tempi recentissimi ha incominciato ad essere recuperato: il *web* è nato solo nel 1990, il primo *browser* grafico e i *plug-in* per il sonoro sono del 1993 (e dunque l'ipertesto multimediale *on-line* nasce in quello stesso anno), le applicazioni *Java* e *Javascript* (che introducono nella Rete la possibilità di 'far girare veri programmi') sono ancora successivi. Il mercato dell'*off-line*, per contro, ha fornito da sempre prodotti *software* di elevata qualità grafica e sonora, con una tendenza a variegare e moltiplicare l'offerta: così, ai ricchi e costosi CD-ROM contenenti enciclopedie multimediali e monografie, si sono aggiunte le pubblicazioni di prodotti a costo contenuto, diffusi attraverso le edicole.

È stato importante rilevare, per sommi capi, le differenze tra testi elettronici *off-line* e *on-line* perché da esse si comprenderanno i diversi gradi di difficoltà che si incontrano nella ricerca dell'informazione su CD-ROM o

¹¹⁰ **Alberto Cadioli** "Le soglie dell'ipertesto", in Bellmann Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.47.

in rete: laddove nel primo caso i problemi sono facilmente risolvibili con la tecnologia disponibile, nel secondo caso spesso essi si mostrano pressoché invalicabili.

3.2. Reperire informazioni off-line

Si è osservato come gli studi delle scienze cognitive abbiano evidenziato l'insufficienza del paradigma della logica classica per comprendere i processi attraverso i quali si sviluppa il ragionamento umano. Sono anche stati presi in considerazione gli studi basati su una specifica attenzione per l'utente finale del prodotto elettronico volti ad indirizzare delle linee guida nel *design* dei documenti. Adesso tenteremo di capire quanti di questi studi abbiano trovato un'applicazione diretta nello sviluppo del *software*, e quanto la tecnologia attuale sia in grado di soddisfare queste linee di ricerca. La risposta, possiamo già anticiparlo, non sarà entusiastica: c'è ancora molto da fare, ma soprattutto il settore tecnologico e quello cognitivo e umanistico devono collaborare maggiormente e non procedere su binari paralleli quando non addirittura divergenti. In particolare cominceremo con il recupero dell'informazione *off-line*, sicuramente più comodo ed efficace del recupero dell'informazione in rete, di cui ci occuperemo in seguito.

Si è già parlato dell'avanzatezza delle tecnologie che utilizzano supporti fisici per la conservazione dell'informazione e di quanto maggiormente avanzate siano rispetto a quelle che utilizzano le interconnessioni tra computer. Nel momento in cui si scrive il CD-ROM è sicuramente lo strumento che fornisce un giusto compromesso tra quantità di informazione memorizzata e capacità di recupero della stessa.

Il CD-ROM (Compact Disc Read Only Memory) nasce nella metà degli anni ottanta da un'idea apparentemente banale: un supporto in grado di memorizzare musica digitale, doveva essere in grado di memorizzare qualsiasi tipo di informazione. È un supporto di memoria ottica che consente la memorizzazione di circa 600 MB di dati, e che deve la sua fortuna alla robustezza, ai bassi costi di produzione e al fatto che le industrie del settore per realizzare i lettori hanno potuto sfruttare gli investimenti già effettuati per il CDDA (Compact Disc Digital Audio). Il suo limite principale è quello di non essere inscrivibile¹¹¹, pertanto si è diffuso soprattutto come supporto

¹¹¹ Questo era vero quando il CD-ROM si è diffuso, una decina d'anni fa; oggi salvare dati su particolari CD registrabili una o più volte (CD-R e CD-RW), o, come si dice in gergo tecnico, masterizzarli, è un'operazione alla portata di tutti.

per enciclopedie, dizionari, *software* di varia natura. Da un certo punto di vista, tale limite si è reso in realtà un punto di forza, poiché il CD-ROM, impedendo i processi di sovrascrittura da parte del lettore, garantisce e preserva la forma testuale progettata dall'autore. La sua nascita ufficiale si fa risalire al 1988, data di pubblicazione della normativa ISO 9660, che ne decretava il formato standard. La velocità di accesso casuale ai dati è effettivamente più scarsa rispetto a quello del disco magnetico, e ciò è dovuto al fatto che la traccia del CD-ROM è a spirale, come quella del cd audio, e non concentrica. Ciò rallenta le operazioni di recupero delle informazioni, perché gli indirizzi sono identificati come 'minuto: secondo: settore' e non dalla combinazione di 'numero traccia' e 'numero di settore', ma aumenta le capacità di memorizzazione, perché si risparmia spazio sul disco. Questa piccola annotazione tecnica ci permette di capire perché il CD-ROM si presta bene all'archivio di informazioni di cui l'elaboratore e l'utente non hanno continuamente bisogno; nel caso di dati di frequente consultazione, al contrario, è più comoda la memorizzazione ('installazione') su disco rigido.

La varietà dei prodotti offerti fa sì che anche la varietà dei sistemi di *information retrieval* sia notevole. Si pensi, infatti, che solo nel 1996 il SIGCAT¹¹² (Special Interest Group on Computer Aided Technology, la più grande organizzazione di utenti di CD-ROM) ha testato dodici diversi 'motori' di recupero del testo (*text retrieval engine*). Si tratta quasi sempre di *full-text retrieval system*, cioè di programmi in grado di ritrovare tra i dati tutte le occorrenze di una parola data. Occorre chiarire che i contenuti multimediali dei CD-ROM ne limitano enormemente le capacità di memorizzazione; tuttavia, l'utilizzo di mappe grafiche, immagini, animazioni, può favorire il processo di autorappresentazione della propria attività di ricerca, a patto che non si riduca a semplice attrattiva accessoria. Le maggiori banche dati in commercio attualmente su CD-ROM fanno un utilizzo spartano della grafica, perché possono contare su numerose chiavi d'accesso all'informazione. ALICE CD, per fare un esempio, è un catalogo di libri in commercio in Italia, consente la ricerca dei dati per autore, traduttore, editore, collana, anno d'edizione, eccetera; ciò è reso possibile dal fatto che una banca dati su CD-ROM, per quanto estesa (350000 libri schedati) è comunque un'area di informazione ben definita, che non necessita di aggiornamento continuo come nel caso di una *web directory*. Questo tipo di sistema di archiviazione bibliografica è estremamente diffuso, a conferma

¹¹² [Http://www.sigcat.org](http://www.sigcat.org)

della sua efficacia: possono essere citati il francese *CD-ROM des livres disponibles*, l'inglese *Global Books in Print Plus*, lo spagnolo *Libros en venta en Hispanoamérica y España Plus*, il tedesco *VLB Aktuell*. C'è uno strumento spesso utilizzato dai curatori di queste banche dati, tradizionale ma efficace: l'*abstract*. L'*abstract* sintetizza in breve i contenuti di un articolo, di un saggio, o di un libro, mantenendo tuttavia un linguaggio tecnico e preciso. Non si tratta, infatti, di riassunti destinati alla divulgazione facilitata dell'articolo; il loro scopo principale è quello di selezionare i contenuti essenziali espressi dall'autore (talvolta l'autore dell'*abstract* e l'autore del testo coincidono) ed evidenziarli in maniera tale che l'utente interessato possa approfondirli, mentre gli altri possano agilmente saltarli. Chi realizza un *abstract* opera un processo di estrazione dell'informazione, che mette a disposizione dell'utente, il quale potrà accettare questa 'offerta', oppure procedere affidandosi al proprio giudizio. Un *abstract*, infatti, non sostituisce in nessun caso il testo originale. Questa stessa strategia di agevolazione del lettore, attuata però attraverso una compressione del testo e non un riassunto, è stata adottata nell'esperienza di cui daremo menzione in seguito, e che a nostro parere si rivela tanto più efficace quanto più utilizza la capacità di scorrimento dell'informazione offerta dai supporti come il CD-ROM. Il processo di produzione degli *abstract* è un lavoro che opera sul piano dei contenuti, richiede l'intervento umano non essendo attualmente automatizzabile con efficacia; in effetti, nelle banche dati su CD-ROM citate in precedenza, solo una parte dell'archivio ne è provvisto.

L'archiviazione di banche dati su CD-ROM è estremamente diffusa anche per quanto riguarda le informazioni di provenienza statale: si fa qui riferimento alla *Bibliografia Nazionale Italiana su CD-ROM*, alla *Bibliographie Nationale Française depuis 1970* al *British National Bibliography on CD-ROM*, al *Deutsche Nationalbibliographie CD-ROM Aktuell*, per citarne solo alcuni. Va sottolineato che tali banche dati raramente utilizzano le tecnologie ipertestuali, preferendo fornire agli utenti un ambiente quanto più possibile simile a quello dell'archivio cartaceo tradizionale, più 'rassicurante' e di facile utilizzo perché permette l'attuazione di tecniche di ricerca ben consolidate. Le banche dati su CD-ROM sono davvero numerosissime, comprendono i cataloghi delle biblioteche (in questi casi, spesso un solo CD-ROM non è sufficiente), cataloghi bibliografici di settori specifici (diritto, economia, sociologia, letteratura, storia, scienze e tecniche, eccetera). Non è davvero possibile citarli tutti, anche perché dal punto di vista che noi privilegiamo ciò non

offrirebbe particolari spunti, oltre a quelli già evidenziati: si tratta cioè di archivi che sfruttano la velocità dell'elaboratore, mantenendo tuttavia le metodologie di ricerca e di catalogazione delle biblioteche cartacee.

Un altro settore, in cui l'archiviazione elettronica si è rapidamente affermata, è quello delle banche dati giornalistiche: i maggiori quotidiani mondiali, infatti, nonché numerose agenzie di stampa, offrono agli utenti i CD-ROM contenenti tutti gli articoli pubblicati, con sistemi di ricerca *full-text*. Ancora più interessanti sono quegli archivi, come il *British Newspaper Index*, che contengono raccolte di articoli forniti di *abstract*.

Il CD-ROM non esclude altri sistemi di memorizzazione di informazione, con i quali, spesso, 'collabora' produttivamente. È il caso per esempio dei CD-ROM che contengono i cataloghi generali di biblioteche o archivi, e forniscono rapidamente la collocazione di un testo cartaceo tra gli scaffali; oppure dei CD-ROM che interagiscono con la rete, fornendo degli indirizzi *web*, per esempio, o permettendo l'aggiornamento attraverso Internet. Ed è il caso dei CD-ROM che, come vedremo nell'analisi dell'esperienza, interagiscono con il libro tradizionale.

A questo punto è possibile delineare una riflessione: la diffusione e il successo di questi supporti elettronici nasce da una corretta suddivisione dei compiti tra operatore umano e operatore elettronico; mentre quest'ultimo, rapidissimo ma fondamentalmente 'stupido', cioè incapace di operare sul piano dei contenuti, permette il recupero quasi immediato di occorrenze espressive, il primo può strutturare l'informazione secondo criteri contenutistici e linee guida improntate sulle necessità dell'utente e le sue possibilità. Non si tratta di strumenti particolarmente rivoluzionari: gli archivi esistono da millenni, e anche l'*abstract*, il riassunto, non è un'operazione progettata e sviluppata in questi anni, benché proprio in questi anni più apprezzata e necessaria che mai. Le nuove tecnologie si mostrano più produttive quando, come ricorda Marchionini¹¹³, non necessitano dell'apprendimento di nuovi saperi e di nuove strategie euristiche, ma permettono e facilitano l'utilizzo degli strumenti già conosciuti dall'utente. Tuttavia, senza lo sviluppo di queste tecnologie non sarebbe stato possibile raggiungere gli stessi risultati per la ricerca dell'informazione. In passato infatti si è messo in pratica il tentativo di fornire percorsi di lettura alternativi di un testo, con il fine di favorire il recupero dell'informazione. In

¹¹³ Gary Marchionini, *Information Seeking in Electronic Environments*, Cambridge, University Press, 1995.

particolare si sono fatti esperimenti sul testo enciclopedico, che meglio si presta ad essere fruito in questa maniera, utilizzando le tradizionali tecnologie della scrittura a stampa: ma i risultati sono stati deludenti. Un bell'esempio a riguardo è descritto da Bolter¹¹⁴: l'Enciclopedia Britannica, nel 1974, in occasione della quindicesima edizione, accompagnò ai volumi dell'opera dotati del tradizionale indice alfabetico, anche un volume, chiamato "Propedia", costituito da un indice per argomenti. Quest'indice rimandava a saggi, articoli dell'enciclopedia a partire dai quali era possibile costruire un percorso di lettura alternativo, trasversale, trasformando così il testo in un vero e proprio ipertesto che il lettore poteva assemblare a piacimento. L'idea era rivoluzionaria, i risultati sono così descritti da Bolter:

In ogni biblioteca che contenesse la *Britannica*, era possibile identificare a colpo d'occhio la *Propedia*, nell'unico volume intonso frammezzo ad altri usati e logori.¹¹⁵

Le qualità specifiche del libro a stampa, la fissità, l'ordine lineare, l'idea che il libro sia un'unità 'chiusa' e indipendente, immodificabile e 'monumentale', hanno impedito l'utilizzo di strumenti di ricerca dell'informazione innovativi: solo il computer, con la sua capacità di generare testi elettronici più 'elastici' ed 'aperti', ha reso possibile tale passo in avanti. Anche in Italia è stato realizzato in passato un prodotto 'proto-ipertestuale' simile: è il caso della Enciclopedia Einaudi (1977-1982), che alla fine di ogni voce tematica presentava i possibili rinvii ad altri passi attraverso una breve sintesi, suggerendo così possibili percorsi nel corpus degli articoli.¹¹⁶

Tornando adesso al discorso fatto per i cataloghi *off-line*, che ne esaltava le capacità di favorire la ricerca immediata dell'informazione, esso muta leggermente direzione se si prendono in considerazione i testi su CD-ROM che sfruttano le possibilità dell'ipertesto multimediale. Quando infatti si permette all'utente di usufruire contemporaneamente di materiale visivo, sonoro, testuale, creando collegamenti ipertestuali, si offre una visione integrata e multidisciplinare dell'oggetto di studi, ma si complica non poco il lavoro di ricerca. Non ritorneremo qui sugli aspetti già trattati in precedenza.

¹¹⁴ Jay David Bolter, op.cit., p. 116- ss.

¹¹⁵ Jay David Bolter, op.cit., p.117.

¹¹⁶ Cfr. Alberto Cadioli, *Il critico navigante [saggio sull'ipertesto e la critica letteraria]*, Genova, Marietti, 1998, p. 103.

Sarà importante ribadire comunque la necessità di fornire all'utente metafore che favoriscano l'adattamento e immagini che favoriscano l'autorappresentazione nonché la rappresentazione del processo. Spiega Barbieri, uno dei responsabili del progetto MuG (Guida Multimediale alla Storia della Civiltà Europea):

La struttura ipertestuale di MuG è stata progettata... in maniera che l'utente possa utilizzare le competenze che già possiede, imparando a usare il sistema senza dover uscire troppo dal quadro concettuale dell'apprendimento storiografico. Lo scopo didattico principale del sistema è quello di fornire un'idea delle cronologie storiche, e dunque della contemporaneità, successioni e distanze temporali tra eventi appartenenti ad ambiti tematici e/o storiografici anche molto diversi tra loro¹¹⁷.

Questi tipi di ipertesti offrono quasi sempre dei percorsi guidati, dal momento che sono stati progettati come strumenti didattici, o informazioni strutturate in maniera tradizionale. Nel caso del progetto MuG, per esempio, esiste uno schedario, una sorta di dizionario enciclopedico che comprende definizioni e descrizioni di rapido accesso, oltre ai collegamenti alle sezioni ipertestuali che sviluppano questi argomenti. Molti prodotti permettono tuttavia il recupero delle informazioni anche attraverso la ricerca di occorrenze testuali. La questione da porre allora, interessante dal punto di vista delle possibilità informatiche che qui stiamo trattando, è: dal momento che questi documenti contengono non solo informazione testuale, ma anche visiva e sonora, come è possibile recuperare quest'ultimo tipo di dati senza ricorrere all'occorrenza testuale? È possibile, insomma, il *multimedia information retrieval*? L'argomento, è bene precisarlo, non riguarda soltanto la ricerca su supporti *off-line*, ma anche su quelli in rete. Tuttavia, dal momento che la tecnologia di questi ultimi è meno avanzata, è facile prevedere che ciò che non è possibile su CD-ROM non è possibile neanche su Internet.

Recuperare immagini "attraverso" immagini può richiedere diverse

¹¹⁷ **Daniele Barbieri**, "Progettare l'interazione", in *Linea Grafica. Rivista internazionale di grafica e comunicazione visiva*, n°283, v.I, Gennaio, cit. in Elena Boriosi Giacanelli, Diana Ascari, 1995, p.196-198.

strategie di programmazione¹¹⁸. La più diffusa è sicuramente quella che si basa sul recupero dell'immagine sulla scorta di informazioni cromatiche. In questo modo la "query" che si pone al programma non è un'occorrenza testuale ma un'immagine campione, scelta per la sua somiglianza con l'oggetto da ricercare. L'elaboratore calcola alcuni valori di questa immagine, come per esempio la luminosità, il contrasto, i colori predominanti, e visualizza le immagini del suo archivio che hanno valori descrittivi simili o uguali. Il sistema può complicarsi notevolmente qualora il programma sia in grado di valutare il punto di vista dell'osservatore, l'orientamento dell'oggetto. I problemi che questi sistemi incontrano sono numerosissimi: in una immagine le unità che non veicolano informazione necessaria, il cosiddetto rumore, sono moltissime, anzi, rappresentano la maggior parte dell'immagine stessa. Si può ricorrere allora al recupero dell'informazione visiva attraverso l'analisi delle forme: una serie di valori numerici sono in grado di rappresentare i contorni di una figura, e di recuperare quindi figure con contorni simili. Accenniamo soltanto all'enorme complicazione che comporta il recupero di informazione da un filmato: ai valori già validi per il recupero di un'immagine statica, bisogna aggiungere i movimenti della macchina da presa, il suo posizionamento, la durata della sequenza, il contesto visivo (ciò che il filmato mostra prima e dopo). Questo campo di ricerche si dimostra comunque promettente, nonostante le enormi difficoltà cui va incontro: il problema di fondo è legato al fatto che mentre per il recupero sul piano espressivo di occorrenze testuali si può contare sulla doppia articolazione del linguaggio, e cioè sul fatto che lettere che compongono gli alfabeti sono di numero finito e si possono quindi facilmente rendere in forma computazionale, ciò non è vero per il codice visivo, che non ha, neanche sul piano espressivo, unità minime sulle quali fare affidamento. Immaginiamo che un utente ricerchi da un *database* visivo di storia dell'arte tutte le immagini (affreschi, quadri, miniature, bassorilievi) che rappresentino la scena della Crocifissione. L'utente potrà inserire come "query" la forma della croce vista frontalmente, e i suoi colori (si potrà per lo meno procedere per esclusione, eliminando le croci arancioni, viola e gialle). Ma in questo caso il computer non sarà in grado di recuperare, tanto per fare un esempio, il *Cristo di San Giovanni della Croce* di Salvador Dalì, che com'è noto raffigura il Cristo visto dall'alto. Insomma, se una ricerca che muova da principi contenutistici per la ricerca testuale è auspicabile, per la

¹¹⁸Cfr. Arnold W.M. Smeulders, Ramesh Jain, *Image Databases and Multi-media Search*. London, World Scientific Publishing, 1997.

ricerca visuale essa diviene quasi indispensabile, viste le difficoltà delle ricerche espressive.

3.3. Reperire informazioni in rete

Quando si pensa alla necessità di ricercare informazioni, si pensa subito ad Internet, o meglio al World Wide Web, la sua interfaccia ipertestuale. Progettato nei laboratori del Cern di Ginevra da Tim Berners-Lee, questo sistema ipertestuale ha fatto della rete delle reti un medium popolare e diffuso, e soprattutto ha aumentato in maniera esponenziale la quantità di documenti elettronici disponibili. E, con essa, la necessità di sistemi di ricerca e reperimento dell'informazione.

Prima di addentrarci nella descrizione degli strumenti attuali, facciamo qualche passo indietro per cercare di comprendere l'evoluzione dei sistemi di ricerca in rete. La data di nascita 'simbolica' della rete che oggi conosciamo si fa risalire di solito al 1957, quando l'Unione Sovietica minacciò il predominio tecnologico e militare degli Stati Uniti lanciando in orbita il primo satellite, lo Sputnik. La reazione statunitense portò alla nascita dell'organizzazione di ricerca Darpa (*Defense Advanced Research Projects Agency*) che avrebbe poi progettato Arpanet, la rete dalla cui evoluzione nascerà Internet. Non c'è accordo fra gli storici sulla presunta origine 'militare' della rete, per cui la rete di calcolatori a commutazione di pacchetto sarebbe servita come strumento di comunicazione alternativo in caso di conflitto nucleare. In ogni caso, non era certo prevista una rete di milioni di documenti in continua crescita, ma piuttosto un sistema di comunicazione resistente ad eventuali crisi locali. A conferma di ciò, basta leggere le parole di Paul Baran, un pioniere della ricerca presso i laboratori della Rand Corporation, il quale nel 1964 afferma che una rete di comunicazione "distribuita", al contrario delle tradizionali reti di comunicazione con un struttura gerarchica, può meglio sopravvivere all'eventualità di un attacco nucleare:

Although one can draw a wide variety of networks, they all factor in two components: centralized (or star) and distributed (or grid or mesh). The centralized network is obviously vulnerable as destruction of a single central node destroys communication between the end stations. In practice, a mixture of star and mesh components is used to form communications networks (...)Such a

network is sometimes called a "decentralized" network, because complete reliance upon a single point is not always required.¹¹⁹

Questo ci permette di affermare senza ombra di dubbio come non si pensasse ad Internet come ad uno strumento di comunicazione per un pubblico di massa, ma piuttosto come ad una rete che collegasse i punti cruciali per la nazione: l'idea di un sistema di *information retrieval* era ancora decisamente acerba, almeno nella mente della maggior parte dei progettisti. Ancora negli anni settanta, infatti, non era in progetto la rete estesa che poi si è sviluppata, come dimostra la struttura dei protocolli di rete TCP/IP, utilizzata a partire dal 1982 e attiva tutt'ora, con uno spazio di indirizzamento relativamente limitato. Ciò che qui si vuole sottolineare è che il carattere 'globale' della rete, la sua diffusione e apertura al vasto pubblico, con la conseguente necessità di realizzare strumenti di *information retrieval* alla portata di tutti, è sorto in tempi piuttosto recenti e in maniera incontrollata, senza che ci fosse un adeguato coinvolgimento degli utenti finali. Persino oggi, denuncia Hauben, il pubblico gioca un ruolo secondario nel processo di pianificazione della rete, come dimostrano le iniziative del governo statunitense di privatizzazione della dorsale della rete attuata senza progettare una vera politica di accesso garantito a tutti.¹²⁰ D'altronde, ricordano gli autori, l'idea di rete di computer nacque dai primi tentativi di *time-sharing*, un'idea che doveva far risparmiare i produttori, più che favorire gli utenti. Le macchine erano troppo costose ed ingombranti per essere utilizzate da una persona sola per volta: nate come 'calcolatori', esse si resero utili come strumenti di comunicazione solo per via indiretta, quando ci si rese conto che in una rete di computer l'interazione tra i programmatori che si scambiavano consigli, dati e impressioni, era molto alta, per cui l'utilizzo del computer stesso come strumento di collegamento tra gli utenti poteva essere una buona idea.

Arpanet (Advanced Research Projects Agency Network) in origine (1969) connetteva soltanto quattro calcolatori, o, più precisamente, quattro reti di calcolatori. Il numero crebbe rapidamente, per cui nel 1971 c'erano 15 nodi e 23 host, quasi tutti laboratori di ricerca universitari (Ucla, Utah, Mit, Harvard, ecc.) ma anche agenzie private (Rand Corporation) e statali (Nasa-

¹¹⁹ Paul Baran, "On distributed Communication Networks", 1964, cit. in Giulio Blasi, op.cit.

¹²⁰ Michael Hauben, Ronda Hauben, *Netizens. On the History and Impact of Usenet and the Internet*. Ieee Computer Society, Los Alamitos, California 1997, p.74.

Ames). Ma probabilmente lo spirito anarchico e democratico della rete attuale, più che ad Arpanet, severamente controllata dallo Stato e dalle origini para-militari, deve molto a Usenet, una rete 'alternativa' che metteva in connessione gli utenti del sistema operativo Unix, sviluppatasi intorno alla fine degli anni settanta. Nata in ambito universitario prima ancora che militare (i primi due nodi furono la Duke University e la University of North Carolina at Chaper Hill), essa si connesse con la 'sorella maggiore' nel 1980 attraverso una sotto-rete, la Berknet, situata al dipartimento di Informatica dell'università di Berkeley, in California, connessa ad entrambe le reti. Mentre Arpanet era una rete fisica dotata di propri cavi ed infrastrutture, Usenet era una rete logica, priva del controllo centralizzato di Arpanet. Ma erano soprattutto le filosofie alla loro base ad essere differenti: alla base di Usenet c'era la semplice e potentissima idea di mettere in comunicazione gli individui, laddove i creatori di Arpanet intendevano fondamentalmente aumentare la produttività dei computer attraverso la condivisione di risorse. Per gli scopi della nostra ricerca, ciò che importa illustrare è che il basso numero di elaboratori presenti in rete venti anni fa rendeva sufficiente, come sistema di catalogazione, un file, *hosts.txt*, che elencava tutti gli *host* e i loro indirizzi IP¹²¹: ogni *host* della rete copiava questo file scaricandolo dal sito in cui era mantenuto ed aggiornato ogni notte.

Bisogna a questo punto fare una precisazione, prima di procedere oltre. Il recupero dell'informazione in rete si scontra fondamentalmente con due problemi, o meglio, con un problema diviso in due fasi. La prima fase consiste nella ricerca del documento che ci interessa; la seconda nel ritrovamento del computer remoto sul quale quel documento è depositato. Mentre la seconda fase è squisitamente informatico-ingegneristica e pertanto non concerne più di tanto la nostra indagine, la prima coinvolge in pieno il nostro approccio linguistico computazionale all'*information retrieval*. Una volta ottenuto l'indirizzo *web* simbolico del sito che ci interessa (www.qualcosa.it), infatti, il nostro computer è in grado di tradurlo in un indirizzo assoluto (124.118.3.7.) ricorrendo ad una procedura di libreria detta risolutrice (*resolver*). Il *resolver* invia un messaggio ad un *server* detto Domain Name Server, una enorme banca dati in continua evoluzione, il quale cerca il nome e restituisce al computer l'indirizzo assoluto corrispondente. A questo punto il contatto tra elaboratori è possibile

¹²¹ Internet Protocol.

attraverso una connessione TCP¹²² con la destinazione.

Benché la questione non sia così semplice come mostrato qui, e benché il problema dei nomi in Internet sia tutt'altro che banale, qui non ce ne occuperemo oltre. Piuttosto, ci soffermeremo sulla prima fase: dove cercare il documento? Il problema infatti sorse subito, già nei tempi in cui i calcolatori connessi alla rete erano di numero abbastanza esiguo. La soluzione che trovarono gli utenti di Usenet per catalogare il materiale in rete di interesse pubblico, che allora era composto soprattutto di *mailing lists e newsgroups*, le liste di discussione alle quali si poteva partecipare dando le proprie informazioni o fornendo le proprie opinioni, fu quella di suddividere le liste per argomento strutturate in maniera gerarchica. Il sistema è davvero molto informale:

Usenet is now made up of thousands of newsgroups organized around different topics. And the number of groups is continually growing. There are procedures for creating new newsgroups in the main hierarchies. When there is a proposal for a new newgroup, those on the Usenet can discuss the proposal. When the vote is called, the new group can be created if there are 100 more votes for the proposed group than against it.¹²³

Tra i primi ambiti di discussione, si ricordano, ovviamente, le scienze dell'informazione, ma anche la fantascienza, la musica, i videogame, gli sport. Argomenti che delineano il ritratto dell'utente dell'epoca, lo studente universitario di informatica, americano, quasi sempre maschio. Arpanet, vale la pena ricordarlo, attraverserà l'oceano atlantico per diventare un po' più rete globale soltanto nel 1982, quando raggiunge il Mathematisch Centrum di Amsterdam.

La progressione del numero di host connessi alla rete tuttavia è in costante crescita. I 213 host connessi nel 1981 saranno 159.000 nel 1989, 17.753.266 nel 1997: la previsione per il 2000 è di 124.255.175 host.¹²⁴ Il balzo in avanti è legato al passaggio da una rete caratterizzata fondamentalmente da messaggistica personale (*e-mail, mailing list, newsgroups*) ad un sistema di

¹²² Transmission Control Protocol.

¹²³ **Michael Hauben , Ronda Hauben**, *Netizens. On the History and Impact of Usenet and the Internet*. Ieee Computer Society, Los Alamitos, California, 1997, p. 61.

¹²⁴ **Giulio Blasi**, op.cit.

informazione ipertestuale. Anche in Italia, seppure con delle enormi difficoltà di natura politica ed economica¹²⁵ che qui non è il caso di prendere in considerazione, la rete sta accrescendo il numero di utenti: tra il dicembre 1997 e lo stesso periodo dell'anno precedente si è passati da 734.000 utenti ad 1.359.000, con un aumento dell'85%.¹²⁶ È il sistema di Berners-Lee la cosiddetta '*killer application*' degli ultimi anni per Internet: il successo del *world wide web* è enorme, perché l'utilizzo di immagini e suoni amplia gli orizzonti creativi possibili, avvicinando a questo mondo utenti senza precedenti esperienze informatiche. Si tenga presente, per comprendere le origini del *web*, che Lee è un fisico che al momento della progettazione operava presso il centro europeo per la ricerca sulla fisica nucleare, laddove numerosi scienziati europei portano avanti ricerche sulla fisica delle particelle. Come ricorda Tanenbaum¹²⁷, il *web* nasce dalla necessità di alcuni gruppi di scienziati di collaborare utilizzando una collezione mutevole di rapporti, schemi, foto, disegni. Il sistema riuscirà perfettamente nell'impresa che invece avevano fallito i suoi 'predecessori', tutti sviluppati tra la fine degli anni ottanta e i primi anni novanta, che oltre ad essere i primi sistemi in grado di fornire un servizio *information retrieval* per gli utenti, erano delle interfacce ancora poco raffinate in verità, degli ambienti attraverso i quali l'utente entrava in contatto con la rete. Nel 1991 appaiono Archie ed Hytelnet: il primo, nella definizione del Foldoc, un dizionario informatico *on-line*¹²⁸, è così definito:

a system to automatically gather, index and serve information on the Internet. The initial implementation of archie by Mcgill University School of Computer Science provided an indexed directory of filenames from all anonymous FTP archives on the Internet. Later version provide other collections of informations.

¹²⁵ Si pensi, tanto per fare un accenno, alla famigerata tariffa urbana, che gli utenti pagano all'azienda telefonica per collegarsi al punto di accesso più vicino alla rete fornito dal *provider*: in Italia questa tariffa è una delle più alte al mondo, con grave ritardo per lo sviluppo dei collegamenti in rete. Ma altri problemi hanno frenato sinora lo sviluppo della rete in Italia: scarsa alfabetizzazione informatica, scarsi investimenti da parte di enti pubblici ma anche aziende private, forte presenza di altri strumenti di comunicazione. La tendenza tuttavia pare finalmente essersi invertita proprio negli ultimi tempi.

¹²⁶ I dati sono citati da **Giuliana Barzon, Patrizia Bertini, Stefania Sordi**, *Trovar lavoro in editoria. Da freelance a imprenditore. Una guida completa alle professioni del libro*. Milano, Editrice Bibliografica, 1999, p.65.

¹²⁷ **Andrew S. Tanenbaum**, *Reti di computer*, Torino, Utet, 1997, p.659.

¹²⁸ FOLDOC, Computing Dictionary, <http://wombat.doc.ic.ac.uk>

Dunque, una directory in grado di archiviare i nomi dei file depositati sui server attraverso il File Transfer Protocol. Per quanto riguarda Hytelnet, esso è definito come segue:

A hypertext database of publicly accessible Internet sites created and maintained by Peter Scott. Hytelnet currently lists over 1400 sites, including Libraries, Campus-Wide Information Systems, Gopher, WAIS, WWW and Freenets.

Non più una semplice *directory*, ma un *database* ipertestuale, che permette l'immediato collegamento con i siti di interesse. Visto che sono stati citati il WAIS e il Gopher, due sistemi del 1991, occupiamoci anche di loro. Il WAIS è

A distributed information retrieval system. WAIS is supported by Apple Computer, Thinking Machines and Dow Jones. Clients are able to retrieve documents using keywords. The search returns a list of documents, ranked according to the frequency of occurrence of the keyword(s) used in the search. The client can retrieve text or multimedia documents stored on the server. WAIS offers simple natural language input, indexed searching for fast retrieval, and a "relevance feedback" mechanism which allows the results of initial searches to influence future searches.

La complessità di questo sistema è notevole, se messa a confronto con quella dei suoi predecessori. Innanzitutto, si tratta di un vero e proprio sistema interattivo, che offre una lista in base ad una richiesta avanzata. Inoltre, è sviluppato in un'ottica volta al soddisfacimento dei bisogni e al linguaggio dell'utente. Ma prima di soffermarci sulle riflessioni d'analisi, chiudiamo la rassegna dei sistemi di *information retrieval* per Internet. Dello stesso anno, il 1991, è il Gopher:

A popular distributed document retrieval system which started as a Campus Wide Information System at the University of Minnesota. Many hosts on the Internet now run Gopher servers which provides a menu of documents. A documents may be a plain text

file, sound, image, submenu or other Gopher Object type. It may be stored on another host or may provide the ability to search through certain files for a given string.

Il Gopher è probabilmente il sistema che più ricorda il *web*, del quale non ha la possibilità di rappresentare immagini: è un sistema esclusivamente testuale. L'anno dopo, il 1992, è il turno di Veronica, un *software* strettamente legato al Gopher, di cui è, se vogliamo, un appendice:

Very Easy Rodent-Oriented Net-wide Index to Computerized Archives. Veronica offers a keyword search of most gopher menu titles in the entire gopher web. As Archie is to FTP archives, Veronica is to Gopherspace. A Veronica search produces a menu of Gopher data source. Because Veronica is accessed through a Gopher Client, it is easy to use, and gives access to all types of data supported by the Gopher protocol.

Ormai la maggior parte di questi sistemi è entrata a far parte, con la velocità che è tipica dello sviluppo tecnologico, della storia dell'informatica. Il *web*, con la sua capacità di presentare interfacce d'interazione amichevoli, ricche di grafica e dall'utilizzo immediato, ha sostituito in buona parte questi sistemi, tranne qualche eccezione: il Gopher, per esempio, è un sistema che riscuote tuttora un certo successo, spesso tra gli utenti cui interessa solo il testo e il recupero in tempi rapidi. Il numero enorme, centinaia di milioni, di pagine in rete, ha messo in crisi il sistema di catalogazione delle *web directories*, costituito da archivi contenenti tutti gli indirizzi ad essi 'sottoposti', cioè suggeriti dagli autori. Gli strumenti attraverso i quali recuperare informazione si sono così automatizzati generando i motori di ricerca. Un motore di ricerca è costituito da tre elementi fondamentali: il *database*, cioè l'insieme degli indirizzi, gli indici e le descrizioni dei siti; l'interfaccia¹²⁹ (la pagina che permette l'interazione con l'utente) e un programma in grado di sviluppare, aggiornare e selezionare gli archivi: il *worm*, il *knowbot*, o lo *spider*.

¹²⁹ Questa interfaccia è andata nel tempo via via ampliandosi, facendo spesso del motore di ricerca un vero e proprio portale, una pagina cioè da cui accedere al *web* che fornisce una serie di indici spesso personalizzabili: è il caso di Yahoo, Lycos, Altavista, tanto per citare le società più note.

I primi sistemi in grado di realizzare automaticamente un indice delle presenze sulla rete sono stati definiti '*worm*', e la loro apparizione si colloca intorno alla fine degli anni ottanta. Il termine *worm* in informatica si riferisce ad un programma che cerca di replicarsi su più *host*: nel caso dell'Internet Worm del 1988, esso era in grado di spostarsi da un sito ad un altro, memorizzandone l'indirizzo; il programma meritava tuttavia qualche perfezionamento, se è vero che mise in crisi migliaia di macchine in rete. Uno di questi sistemi attualmente ancora attivo è il WebCrawler, anche se oggi si parla più spesso di *knowbot*, anziché di *worm*. Lo spostamento terminologico non è irrilevante: il tentativo dei programmatori è stato quello di trasformare un programma di catalogazione fondamentalmente incontrollabile in qualcosa di più 'intelligente', un agente in grado di effettuare ricerche per conto dell'utente sulla base di alcune informazioni ricevute. Il programma è così definito dal suo autore Brian Pinkerton:

The WebCrawler is a web robot, and is the first product of an experiment in information discovery on the web. I wrote it because I could never find information when I wanted and because I don't have time to follow endless links.¹³⁰

Il WebCrawler si comporta come un agente in grado di ricercare documenti di particolare interesse per l'utente: dapprima, l'interfaccia gli offre un indice di quanto è già stato catalogato, strutturato per contenuti; in seguito, è in grado di interagire con l'utente sulla base di alcune richieste (*query*) di quest'ultimo, per indirizzarsi verso un particolare tipo di documento. La catalogazione dei documenti avviene attraverso i *link*: partendo da un documento, il programma ne individua i collegamenti ad altri documenti sui quali si sposta. La strategia di base di questo programma è quella di navigare la rete in maniera piuttosto 'superficiale', costruendo una mappa dei siti presenti partendo da quelli più spesso visitati: solo nel momento in cui c'è una richiesta specifica dell'utente, si restringe il campo e si sviluppa la ricerca 'in profondità'. Un ulteriore modello di sistema di recupero dell'informazione sul *web* è il cosiddetto *spider*, un programma che oltre a cercare informazioni per costruire enormi *database* testuali, è in grado di conservare dei siti *web* sulla base delle richieste fatte da chi lo gestisce,

¹³⁰ Sandy Ressler, *The Art of Internet Publishing, The Internet and Beyond*, Upper Saddle River, Prentice Hall, 1997, p.21.

cercando di individuare le logiche che governano le relazioni tra siti, eliminando i siti non più attivi, aggiungendo i nuovi arrivi.

Ricapitoliamo in breve il funzionamento di questi sistemi (genericamente, *robot*). La struttura del *web*, ridotto alla sua parte essenziale, è piuttosto banale: un grosso grafo in cui i nodi sono le pagine e gli archi sono i puntatori o *link*. Il motore opera dapprima una fase di ricerca, elaborando le Url¹³¹ delle pagine che incontra e controllando che esse non siano già state memorizzate: se così non è, la pagina è recuperata, gli Url e i titoli vengono copiati, e i puntatori forniscono nuovi indirizzi dai quali far ricominciare il ciclo. Questi puntatori non vengono seguiti nell'ordine ottenuto, perché ciò comporterebbe una ricerca in profondità, a scandaglio, tante volte quanto è lungo il massimo cammino non ciclico. Invece i motori di ricerca operano a ventaglio: prima memorizzano tutti i *link* di una pagina in un vettore, poi passano alla pagina immediatamente successiva memorizzata dal vettore stesso.

Una volta effettuata la ricerca, è necessario indicizzare le parole chiave: questo passaggio è importante per scartare le preposizioni, le congiunzioni, gli articoli, e tutte le parole così comuni da essere di scarso valore semantico. Le parole chiave, che come abbiamo detto quasi sempre sono quelle del titolo, sono poi affiancate all'indirizzo della pagina che le contengono, e così memorizzate sul disco. Da questo disco il programma cercherà le parole richieste dall'utente e fornirà tutti gli Url collegati. Esiste la possibilità, da parte di chi gestisce un sito, di impedire l'accesso a questi 'visitatori', attraverso un protocollo, il *Robot Exclusion Protocol*: prima di poter catalogare un sito, il robot deve controllare in un registro che esso non sia 'interdetto'. Ma è il caso opposto ad essere più frequente, quello cioè che l'ideatore di un sito voglia farsi catalogare dai motori di ricerca: in questo caso deve procedere ad una 'sottomissione' manuale, oppure affidarsi a *software* particolari la cui funzione è proprio quella di segnalare un sito ai motori di ricerca.

I *robot* spesso richiedono sistemi operativi avanzati e complessi per funzionare, come lo Unix, oppure macchine piuttosto costose, *workstations* di dimensioni notevoli; in ogni caso, consumano spesso molta banda con la

¹³¹ Uniform Resource Locator: è la carta d'identità di documento di Internet, che ne individua il protocollo (http nel caso delle pagine *web*, mailto per la posta elettronica, news per i gruppi di discussione, eccetera) il nome DNS della macchina sulla quale si trova la pagina, e il nome del file che indica la pagina specifica.

loro presenza ingombrante. Perciò l'utente delle rete lascia che a fare questo lavoro di catalogazione siano delle società, che poi offrono gratuitamente (o quasi, vista la presenza di *banner* pubblicitari: ma è il minimo per garantire la sopravvivenza di un servizio 'libero') i risultati delle loro ricerche: i cosiddetti *Internet Starting Points*, o *Jumpstations*, o *Net Directorites*. Le modalità con cui questi sistemi funzionano nel raccogliere informazioni e fornirle al pubblico sono piuttosto variegata. Solo alcuni sono in grado di indicizzare tutte le parole che trovano nei testi delle pagine *web* attraverso l'uso di programmi sofisticati (i vari *worm* di cui dicevamo prima); altri continuano ad operare esclusivamente come *web directory*, dunque sono gestiti da esseri umani che creano archivi sulla base di ciò che selezionano in prima persona o ciò che gli viene consigliato dagli utenti stessi; alcuni presentano un indice strutturato per contenuti dei loro database.

Sebbene non sia possibile un'indagine accurata di tutti i motori di ricerca (sono centinaia) è doveroso tuttavia delinearne le strategie di ricerca adottate. Abbiamo visto che la maggior parte di loro raccoglie le parole chiave con le quali classificare il sito nel suo titolo, e quindi all'interno dei due *tag*¹³² HTML <TITLE> contenuti nell'intestazione del documento. Esistono però altre possibilità¹³³: l'autore del documento *web*, infatti, può inserire ulteriori informazioni sul suo sito all'interno del tag <META>, e in particolare suggerire direttamente alcune parole chiave (non c'è un limite al loro numero, ma la maggior parte dei robot considera solo le prime) e fornire una breve descrizione. Si tratta in questo caso di quella frase che di solito accompagna un indirizzo in un indice: se manca, il motore di ricerca prende la prima frase del testo. Lo scrittore ipertestuale può fare uso inoltre di titoli particolarmente evidenti (gli *heading paragraph*) nella consapevolezza che probabilmente il robot li indicizzerà.

L'elemento comune a tutti i motori di ricerca, e su questo punto baseremo la nostra riflessione, è dato dal fatto di gestire la loro interazione con l'utente sulla base di *keyword*, parole chiave, strutturate attraverso connettivi logici rudimentali.

3.4. I limiti dei motori di ricerca

Scrivono Smoliar, Baker, Nakayama e Wilcox, in un loro saggio molto

¹³² Un *tag* è una annotazione dell'HTML, l'HyperText Mark-up Language, della cui struttura si parlerà in seguito.

¹³³ Cfr. *Il sito è...Sottoposto*, in Internet News n°5, maggio 1999, p.58-68.

interessante:

The better part of this decade has seen considerable effort towards advancing technology in "content-based image retrieval"(...) Unfortunately, none of this work really has anything to do with content in Hjelmslev's sense of the word. A humbler awareness of history and a more generous acknowledgement of Hjelmslev's perspective would rightfully call this technology *expression-based retrieval*¹³⁴.

Gli autori in questo periodo si riferiscono ai sistemi di ricerca di informazioni multimediali, e in particolare di immagini, tuttavia la loro conclusione può essere generalizzata a tutti i sistemi di ricerca automatizzata, perché ne delineano la caratteristica principale: l'essere capace di effettuare ricerche solo ed esclusivamente sul piano dell'espressione. Nello stesso saggio, infatti, si sostiene che i motori di ricerca attuali possono offrire una selezione operata solo sul piano dell'espressione, lasciando poi all'utente l'analisi contenutistica dei vari testi suggeriti e quindi una ulteriore definitiva selezione.

Questo concetto, per quanto banale possa apparire, è di importanza vitale. I sistemi di *information retrieval* automatizzati si muovono solo ed esclusivamente sul piano dell'espressione. Le attuali possibilità tecnologiche fanno sì che il sistema riceva una stringa di caratteri ASCII dall'utente, traducendola in un codice numerico, la cui sostanza dell'espressione è, come noto, una serie di impulsi elettrici. A questo punto il programma va alla ricerca nei propri archivi dello stesso codice e ripropone i risultati. Questo sistema pone non pochi limiti alla ricerca delle informazioni.

Innanzitutto, un problema banale, è che il programma non è dotato di un meccanismo intelligente in grado di riconoscere l'errore ortografico: se una parola è digitata male, non verrà trovato alcun documento corretto. Può sembrare una sciocchezza, ma non lo è, se si considera la tesi esposta in precedenza per cui bisogna far compiere al computer le operazioni in cui è più capace della mente umana. Un utente non è tenuto a memorizzare centinaia di nomi, si pensi ai nomi propri in qualche lingua straniera

¹³⁴ Stephen W. Smoliar, James D. Baker, Takehiro Nakayama, Lynn Wilcox, "Multi-Media Search: An Authoring Perspective", in A.W.M. Smeulders, R. Jain, 1997, op.cit., p.7.

particolarmente ostica: basta dimenticare una lettera per vanificare lo sforzo. Certo, si potrebbe obiettare, si può sempre cercare un documento, cartaceo o elettronico, che contenga la corretta scrittura del nome, e riproporre la ricerca. Ma un ragionamento di questo tipo è un ragionamento di chi 'si accontenta', si adatta, modella l'uomo alle possibilità della macchina, mentre il principio dovrebbe essere opposto. Con ogni probabilità, in un futuro prossimo i sistemi di ricerca saranno in grado di suggerire alcune diverse scritture di un nome, così come oggi alcuni *word-processor* suggeriscono correzioni ortografiche, riconoscendo alcune delle lettere e confrontandole con gli archivi di cui sono attrezzati.

Un secondo, e anche in questo caso risolvibile problema, è che il sistema potrebbe non ritrovare una Url richiesta, perché obsoleta. In questo caso, ci sono due possibilità: la prima è che il *link* suggerito dal motore di ricerca punti ad una pagina che non esiste più o che è stata spostata; e in questo caso il *server* su cui dovrebbe trovarsi risponde con un codice di errore, generando il famelico 'Url not found'. La seconda possibilità è che il *server* cui il puntatore è rivolto sia spento, o fuori uso: in questo caso potrebbe esserci un ritardo anche notevole dell'attesa. La soluzione è impostare un *timeout*, una temporizzazione che interrompa il lavoro di ricerca del *browser* dopo un tempo definito. Il rischio in questo caso è che il *timeout* sia troppo breve e scarti Url valide, oppure troppo lungo, causando notevoli perdite di tempo.

Un terzo problema coinvolge l'aspetto prettamente multimediale del *web*. Abbiamo visto che i *knowbot* e i *worm* indicizzano il materiale partendo dai titoli, dai tag metatestuali, o, in qualche caso, memorizzando l'intero testo. E se il *link* non è testuale? In rete è molto diffuso l'uso di mappe grafiche per spostarsi da una pagina all'altra. In questo caso, non tutti i motori di ricerca sono attrezzati per garantire il ritrovamento della pagina. Tuttavia, con una buona accortezza da parte del realizzatore della pagina, anche questo problema può essere risolto, fornendo sempre dei *link* alternativi alle mappe. Lo stesso identico problema è generato dai *frame*, pagine costituite da strutture prive di informazione, che contengono altre pagine; dal *frame* il robot non può ricavare informazione indicizzabile. In questo caso il *webmaster* può scrivere una pagina senza *frame*, ad uso esclusivo del *robot*, per fornirgli le occorrenze testuali da indicizzare.

Ed è sempre l'accortezza del *webmaster* che può evitare il quarto problema della nostra serie: quello delle pagine che non hanno un titolo significativo, o che non ce l'hanno proprio. Abbiamo visto che il sistema

ricerca l'occorrenza suggerita solo tra le parole che ha già registrate nei suoi archivi. Se si tratta di un motore di ricerca come, per esempio, Open Text Index, il programma avrà in archivio tutte le parole dei siti *web* catalogati, e il problema del titolo non si pone. Ma per la maggior parte dei motori, la parola chiave andrà cercata tra i titoli dei documenti. Questo riduce enormemente la ricerca, ma anche le possibilità di ritrovare il documento ricercato: se l'autore di un documento *web* non è stato accorto nella scelta del titolo del documento, difficilmente il suo lavoro potrà essere recuperato. Se invece il documento non ha affatto un titolo (e sono moltissimi), la responsabilità dell'autore è completa, e non si può accusare l'elettronica: neanche un libro cartaceo è rinvenibile senza elementi paratestuali fondamentali come il titolo. Anche in questo caso, il problema è diversamente affrontato dai gestori dei vari motori di ricerca, e qui non possiamo certo prenderli in considerazione tutti. Una buona soluzione, adottata da alcuni, cui ci limitiamo a fare qualche accenno, è quella di indicizzare solo le parole 'importanti'. Si tratta di una scelta a metà strada tra l'indicizzazione esclusiva dei titoli, che appare un po' troppo limitativa, e quella di tutto il testo, che aumenta in maniera esponenziale la dimensione del *database*, e i tempi di reperimento. Quali sono le parole 'importanti'? Il criterio, com'è ovvio, non può che essere quantitativo: si indicizzano le parole con il maggior numero di occorrenze, salvo quelle 'escluse a priori' come gli articoli o le preposizioni. È una soluzione perfettibile, è ovvio. Basti pensare che se l'autore di un sito di critica dantesca vuole che il suo testo sia rinvenuto da questo sistema, dovrà ripetere il nome Dante quante più volte possibile, evitando pronomi e perifrasi, con buona pace della sua coscienza di italianista.

La 'stupidità' dei *robot*, il loro muoversi automaticamente in un piano espressivo alla ricerca di occorrenze senza poterle minimamente interpretare, è all'origine di una serie di trucchi, piuttosto disonesti, da parte degli autori ipertestuali per dare visibilità al proprio sito, o, per usare un termine gergale molto diffuso, per migliorare il proprio *ranking*¹³⁵. Uno è quello di cambiare continuamente titolo alla propria pagina, generando così più registrazioni nei motori di ricerca: sistema che crea numerosi doppioni e non facilita certo la ricerca dell'informazione. Un altro trucco consiste nel generare una pagina (*lead page*) che ripete ossessivamente due o tre parole chiave, scalando di conseguenza la graduatoria dei motori di ricerca, e che rimanda poi al sito

¹³⁵ In inglese "rank" significa letteralmente "fila, riga", ma anche "ceto sociale". Il termine è stato pertanto adottato nel gergo della rete, indicando la posizione occupata da un sito all'interno di un motore di ricerca: essere tra le prime dieci, venti posizioni è il desiderio di tutti i realizzatori delle pagine *web*.

vero e proprio. Ancora più discutibile eticamente è la creazione di un *auto forwarding page*, una pagina che oltre a contenere numerose occorrenze della parola chiave, contiene un codice per rinviare direttamente al sito vero e proprio: una tecnica che molti motori di ricerca ostacolano, eliminando gli indirizzi che la praticano. Per alcuni motori, come abbiamo già detto, l'unica strada percorribile per indicizzare il proprio lavoro è quella di suggerire direttamente sia l'indirizzo del documento da inserire nell'archivio, sia una serie di parole chiave determinanti per la ricerca del sito. È un sistema certamente più lento, ma che permette un'opera di filtro editoriale dal valore non indifferente. È il caso di Yahoo, l'unico motore di ricerca di dimensioni notevoli a seguire una politica decisamente contenutistica: i siti non vengono automaticamente indicizzati dai robot, ma classificati da operatori umani uno per uno nelle categorie di appartenenza. L'operazione rallenta l'aggiornamento del sistema, a tutto vantaggio della qualità, che ne fa a giusta ragione il più popolare al mondo. In ogni caso, neanche la selezione umana può risolvere il serio problema della vastità di riscontri recuperati da un motore di ricerca, perché nessun ricercatore, per quanto paziente, può visionare qualche migliaio di siti alla ricerca di quello interessante, con il rischio di non ritrovarlo. La soluzione adottata per risolvere questo problema è stata quella di restringere progressivamente il campo delle ricerche¹³⁶. Alcuni, come Altavista¹³⁷ per esempio, offrono la possibilità di scegliere il linguaggio del sito da recuperare: sicuramente un bel vantaggio. Ma le possibilità di restringere il campo di ricerca si muovono soprattutto seguendo la logica dei connettivi booleani. Si può cioè suggerire al motore di ricerca di elencare tutti i gli indirizzi dei documenti che contengono sia una parola che una seconda parola data, attraverso il connettivo AND; oppure quelli che ne contengono almeno una, secondo il connettivo OR. Nel primo caso il sistema opererà un'intersezione degli insiemi dei risultati ottenuti, nel secondo caso un'unione. Ma l'utilizzo dei connettivi non si limita a questi due. Si può usare infatti anche il connettivo NOT, che esclude dalla ricerca i documenti contenenti una determinata parola, oppure NEAR, che richiede la presenza 'ravvicinata' di due termini. I criteri per stabilire tale vicinanza variano: 25 parole per Lycos¹³⁸, ad esempio, 10 per Altavista. Il principio può essere capovolto, cercando cioè documenti che contengono due parole a patto che siano lontane, o 'esasperato', quando le parole da rinvenire devono essere

¹³⁶ Cfr. **M.Valente**, *Motori di ricerca alla massima potenza*, in "Internet News" n°2 febbraio 1999, p.32-35.

¹³⁷ <http://www.altavista.digital.com>

¹³⁸ <http://www.lycos.com>

adiacenti. Esiste inoltre la possibilità di estendere la ricerca a tutte le parole contenenti un radicale, ed infine la possibilità, piuttosto interessante, di operare una ricerca selezionando i documenti a seconda dei loro contenuti multimediali (suoni, immagini, javascript, java, Active X, VRML, eccetera). Infine, racchiudendo le parole tra virgolette, si permette al motore la ricerca di un documento che contiene esattamente quella sequenza di caratteri.

Pur essendo stati fatti notevoli passi avanti, il problema è ancora lungi dall'essere risolto. Se, come accennato in precedenza, la logica classica non è più considerata sufficiente per formalizzare il linguaggio umano, figuriamoci una logica fatta di pochi connettivi. Facciamo un esempio: si ponga il caso che l'utente voglia cercare un saggio di un certo Mario Rossi pubblicato su una certa rivista elettronica mensile di cui non ricorda il nome. La sua richiesta potrebbe essere all'incirca la seguente:

rivista AND ("Mario Rossi" OR "Rossi Mario") NOT settimanale NEAR rivista.

Più che una richiesta in linguaggio naturale sembra un'operazione algebrica, e non è che un esempio piuttosto semplice: è l'utente che si adatta al linguaggio della macchina, e non il contrario. A proposito delle ricerche attraverso connettivi, Boriosi e Ascari, mostrando una scheda rappresentante una ricerca su banca dati effettuata attraverso gli operatori booleani, affermano:

Dalla scheda si può ancora una volta notare l'estrema precisione con cui il computer può condurre una ricerca ma anche il fatto che questa precisione è completamente priva di intelligenza (almeno nei computer normali che non sono ancora dei cosiddetti *Sistemi Esperti*), cioè della capacità di scelta in base ai risultati. Il computer quindi non riconosce gli errori a meno di non indicarglieli. È quindi bene fissarsi in mente l'idea che una ricerca al computer è sempre una ricerca assistita dal computer che deve essere diretta da un'intelligenza umana.¹³⁹

Quest'ultima frase è molto importante: i computer attuali sono in grado di

¹³⁹ **Elena Boriosi Giacaneli, Diana Ascari**, *Guida alle ricerche bibliografiche. Dalla biblioteca alle banche dati alle reti telematiche*. Bologna, Zanichelli, 1995, p.163.

assistere l'intelligenza umana, non di sostituirla, almeno per quanto riguarda l'*information retrieval*. Essi richiedono la conoscenza di un sistema di regole formalizzato che, come abbiamo visto, non è proprio semplicissimo; attualmente non sono in grado di riprodurre un'interazione basata sulle modalità del ragionamento comune. Abbiamo detto in precedenza che il nostro modo di ragionare segue cammini più circolari che lineari, con una certa ricorsività di alcuni percorsi; il passaggio tra dati, ipotesi e tesi non è un univoco, ma segue direzioni di percorrenza variabili. La ricerca su rete attuale permette una sola direzione di ricerca; non si può ritornare su un passaggio precedente che si è rivelato errato. Un ulteriore, grave limite attuale è che la ricerca in rete non permette la restrizione 'graduale' dell'area di ricerca; tutte le volte che si pone una *query*, il motore inizia una ricerca completa tra tutti i suoi archivi, e non, come sarebbe più utile, soltanto tra i risultati della ricerca più recente. Questo limite sta tuttavia per essere superato: alcuni motori di ricerca¹⁴⁰ permettono già oggi una riduzione graduale del numero di riscontri, attraverso una serie di selezioni consecutive, per cui si possono scartare volta per volta centinaia di occorrenze non rilevanti. Non si smentisce dunque l'assunto semiotico per cui ampliando l'intensione di un termine se ne riduce l'estensione.

Un altro tentativo di risolvere il problema dell'eccessiva quantità di indirizzi raccolti dal motore di ricerca è ravvisabile nella nascita di meta-motori di ricerca: motori che cercano non direttamente in rete, ma negli archivi degli altri motori di ricerca. Tuttavia, da un punto di vista linguistico computazionale, il problema non si risolve aumentando la gamma dei connettivi, o operando una restrizione di campo piuttosto casuale e non basata su principi qualitativi. Alla base delle difficoltà dei motori di ricerca c'è il fatto, lo ripetiamo a costo di apparire noiosi, che la macchina ricerca solo occorrenze espressive, disinteressandosi del contenuto. Se, per fare un esempio banale, ponessimo una ricerca della parola 'gamba', non troveremmo un documento che si chiama "La struttura degli arti inferiori", e magari troveremmo un sito che parla delle (solite) gambe dei tavoli. L'esempio è banale, il problema talvolta drammatico, se è vero che ormai quella del 'ricercatore *web*' è una vera e propria professione di cui solo le grosse imprese possono disporre. Occorre, secondo il nostro parere, un ritorno al contenuto, alla semantica, troppo spesso messa da parte da chi

¹⁴⁰ Per esempio, in Italia, il servizio offerto all'indirizzo <http://www.kataweb.it> permette un tale tipo di ricerca. Probabilmente questo sistema, molto efficace, è destinato in tempi brevi a diffondersi per tutti i motori di ricerca.

opera nel settore tecnologico. E poiché, almeno finora, i risultati dell'Intelligenza Artificiale sugli aspetti semantici dell'elaborazione del linguaggio sono stati insufficienti, bisognerà che ad occuparsi di contenuti sia l'intelligenza umana. È necessario cioè pensare un'operazione di sintesi, filtro, compressione testuale svolta a monte, perché il problema non potrà che peggiorare in futuro. I maggiori operatori nella progettistica dei motori di ricerca se ne sono accorti da tempo. Si è già detto che per entrare negli archivi di alcuni portali, i siti di accesso ad Internet organizzati da alcune società, occorre aver subito una selezione gestita da operatori umani che eliminano le pagine non rilevanti per il proprio pubblico. Dell'intelligenza umana, della capacità di valutare qualitativamente un documento, insomma, per il momento non si può fare a meno.

Ed è questo aspetto che ci permette di trarre una conclusione importante che riprenderemo oltre: la complessità dei processi di *information retrieval*, e l'enorme mole di lavoro che essa comporta, può essere semplificata attuando a monte, da parte dell'elaboratore del testo, alcune operazioni selettive, che saranno quindi 'risparmiate' all'utente finale. Abbiamo visto che questo è il caso di quei motori di ricerca che classificano il materiale da loro ricercato per categorie, dopo un'operazione di giudizio semantico gestita da operatori umani. La selezione del materiale più adeguato e lo scarto di quello più approssimativo, operazione di cernita che ancora una volta solo degli operatori umani in grado di trattare qualitativamente l'informazione possono eseguire, evita al navigatore di imbattersi in siti poco professionali (a meno che questa non sia la precisa volontà del navigatore stesso). Anticipiamo qui che l'esperienza che descriveremo in seguito è stata caratterizzata proprio dallo svolgimento di un processo di estrazione dell'informazione da parte dei curatori che hanno inteso così 'semplificare' la successiva attività cognitiva del lettore, prendendosi carico di una parte del processo.

Volendo chiudere questa sezione 'informatica' con una prospettiva rivolta agli orizzonti futuri, pare opportuno suggerire un orizzonte 'di speranza', e uno decisamente più cupo. Il primo si fonda sulla possibilità di compiere un passo in avanti importante: permettere alla macchina di riconoscere non solo delle singole parole, come attualmente avviene, ma delle unità espressive più ampie, e semanticamente connotate: introduzioni, riflessioni, commenti, citazioni. Il calcolatore non è in grado, ovviamente, di riconoscere le diverse funzioni che i periodi espressivi svolgono nell'ottica dell'elaborato testuale.

A meno che non sia l'autore stesso a fornire queste informazioni, attraverso i sistemi di *mark-up*, in grado di conciliare la rapidità di ricerca delle macchine con la competenza semantica degli esseri umani. Il secondo, meno rincuorante, riguarda il problema del recupero dell'informazione per i ricercatori dei secoli futuri. Sarà doveroso infine accennare alle prospettive offerte dalla ricerca in Intelligenza Artificiale relative allo sviluppo di programmi 'intelligenti', i cosiddetti sistemi esperti.

3.5. Il mark-up e le possibilità offerte all'information retrieval

Abbiamo già citato la critica che Coombs, Renear e DeRose muovevano più di dieci anni fa all'indirizzo della ricerca scientifica che si preoccupava soprattutto di fornire macchine sempre più veloci, senza approfondire la possibilità che le stesse contribuissero al miglioramento del processo di elaborazione e ricerca dell'informazione:

We see far more attention paid to keyboards, printers, fonts, displays, graphics, colors, and similiar features than to the retrieval and structuring of information or even to the verification of spelling and grammar¹⁴¹.

Per quanto riguarda l'ausilio grammaticale, indubbiamente sono stati fatti dei progressi notevoli. Per quanto concerne invece la strutturazione e il recupero dell'informazione, le proposte degli autori sono rimaste tuttora scarsamente realizzate. Non si sfrutta appieno il fatto che il testo elettronico è caratterizzato da "l'oscillazione rapida tra la dimensione della trasparenza e quella dell'opacità"¹⁴²: cioè la sua natura permette di guardare al testo superficiale, ma anche 'attraverso il testo', rinvenendone la struttura immanente. Questo processo di guardare 'attraverso' può favorire notevolmente il recupero dell'informazione, attraverso l'utilizzo di strumenti come i cosiddetti linguaggi di *mark-up*.

La traduzione italiana del termine potrebbe essere quella di 'annotazioni', anche se non rende perfettamente l'idea. Annotazioni che l'autore del testo pone accanto allo stesso per descrivere le modalità di formattazione del testo. Il sistema di *mark-up* più noto è senz'altro l'HTML, lo standard di formattazione per le pagine *web*, che in maniera imprecisa è considerato un

¹⁴¹ James H.Coombs, Allen H.Renear e Steven J.DeRose, op.cit., p. 934.

¹⁴² Jay David Bolter, op.cit. p.214.

linguaggio, essendo una specifica, un tipo del più generico linguaggio SGML¹⁴³, o, per usare un termine tecnico, un DTD (document type definition). Per fare un esempio di *mark-up*, l'annotazione (*tag*) <bold> dell'HTML indica al programma che visualizzerà il documento che il testo compreso tra i tag dovrà essere visualizzato in neretto. Tuttavia il *mark-up* non si limita a fornire delle indicazioni sul piano dell'espressione testuale, ma anche su quello contenutistico: il tag <cite>, per esempio, indica che il testo che segue è una citazione. L'utilizzo dei sistemi di *mark-up* per la codifica dei testi è, secondo Gigliozi, l'iniziativa maggiormente foriera di risultati tra le applicazioni umanistiche dell'informatica, viste le possibilità che offre non solo al trattamento, la conservazione e la trasmissione dell'informazione, ma anche all'analisi degli elementi che compongono un testo e dei rapporti che intercorrono tra questi elementi¹⁴⁴. Prima di approfondire il discorso, è opportuno chiarire alcune distinzioni sui vari sistemi di *mark-up*.

Il *mark-up* può essere specifico o generico. Nel primo caso, esso dipende fortemente dal sistema che ha generato il documento, poiché definisce una formattazione valida esclusivamente per il programma usato. È il caso della maggior parte dei *word-processor*, che formattano il testo in maniera tale che sia visualizzabile solo da quel programma, o comunque da un programma dotato di supporto di 'traduzione'. Si pensi alle difficoltà che si riscontrano nell'aprire e visualizzare perfettamente un documento Microsoft Word, con un diverso programma, per esempio WordPerfect della Corel. Il *mark-up* generico, al contrario, fa della portabilità (termine che in informatica indica la possibilità di essere utilizzato su macchine diverse) la sua forza. Il *mark-up* generico, infatti, è completamente indipendente dal sistema che genera i documenti: esso fornisce informazioni contenutistiche (questa è una citazione, questa è una frase rilevante) lasciando poi al *software* di visualizzazione il compito di stabilire come tradurre queste informazioni contenutistiche in informazioni grafiche e stilistiche.

¹⁴³ Standard Generalized Mark-up Language. È uno standard Iso pubblicato nel 1986 e modificato nel 1988. Deriva dalle esperienze congiunte di Goldfarb con il GML (primi anni settanta) e di Reid con il sistema Scribe (Carnegie-Mellon, fine anni settanta). Da un punto di vista strettamente semiotico, definire SGML un linguaggio è un errore: si tratta infatti di un metalinguaggio, in quanto consiste in una serie di norme sintattiche astratte, attraverso le quali è possibile definire infiniti sistemi di codifica. Cfr. **Sandy Ressler**, *The Art of Internet Publishing, The Internet and Beyond*, Upper Saddle River, Prentice Hall, 1997.

¹⁴⁴ **Giuseppe Gigliozi**, "Critica letteraria e nuove tecnologie", in Bellmann Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.63.

Tuttavia questa suddivisione è piuttosto generica e merita di essere approfondita. I sistemi di *mark-up*, infatti, possono essere catalogati in sei categorie¹⁴⁵: puntuazionale, presentazionale, procedurale, referenziale, metamark-up, descrittivo.

Il *mark-up* puntuazionale consiste nell'uso di annotazioni che forniscono informazioni prevalentemente sintattiche riguardanti il testo. Questo genere di *mark-up* fonda il suo funzionamento sull'interpretazione dell'uso della punteggiatura da parte dell'autore, e di conseguenza soffre di notevoli limiti. L'uso della punteggiatura, infatti, è tutt'altro che oggettivo: l'utilizzo della virgola, o del trattino, possono variare a seconda delle scelte stilistiche dell'autore; non è detto che un punto indichi la fine di una frase: potrebbe trattarsi per esempio di un'abbreviazione. Inoltre questo sistema genera spesso delle ambiguità, per la sua scarsa capacità di interpretare il senso dello scritto basandosi solo sulla punteggiatura, e richiede spesso controlli successivi. Il suo unico vantaggio è quello di richiedere poco lavoro da parte dello scrittore, che si limita ad usare il sistema di punteggiatura che avrebbe utilizzato comunque, senza dover compiere ulteriori attività descrittive.

Il *mark-up* presentazionale si limita, come dice il termine stesso, alla presentazione grafica e visiva del testo. Indica insomma le interruzioni di pagina, l'uso degli spazi, la posizione delle note, eccetera. È un sistema decisamente obsoleto, perché quasi tutti i programmi di scrittura operano queste annotazioni automaticamente, e sono infatti definiti WYSIWYG¹⁴⁶ perché permettono la visualizzazione del documento e contemporaneamente l'apporto di modifiche.

Il *mark-up procedurale* non descrive le caratteristiche sintattiche o grafiche del documento, ma indica all'elaboratore alcune procedure precise da compiere: salta una riga, cambia carattere, inizia una nuova pagina. Il *mark-up procedurale* è tecnicamente molto efficiente, ma è così strettamente legato alla macchina di elaborazione da essere da alcuni autori¹⁴⁷ completamente identificato con il *mark-up* specifico. Ciò nonostante è il più diffuso, unito al *mark-up* presentazionale: il *software* cioè interpreta l'aspetto grafico del testo elaborato dall'utente, e lo traduce automaticamente in una serie di procedure da attuare. Il Word di Microsoft, per esempio, è in

¹⁴⁵ Si tratta di una suddivisione intrapresa qui per comodità di esposizione e tutt'altro che definitiva: i testi a riguardo spesso riducono le categorie a tre, oppure le definiscono con altri nomi. Aldilà delle scelte classificatorie, i concetti di base, tuttavia, sono sempre gli stessi.

¹⁴⁶ What You See Is What You Get

¹⁴⁷ Sandy Ressler, op.cit., p. 168.

grado di distinguere un paragrafo da un altro, attraverso l'utilizzo, da parte dell'utente, del tasto "invio", che genera uno spazio 'grafico'.

Il *mark-up* referenziale 'fa riferimento' ad entità esterne al documento che poi saranno sostituite durante l'elaborazione. Per esempio si utilizza una abbreviazione che fa riferimento ad un termine memorizzato in un altro *file* dall'elaboratore, che in seguito provvederà alla sostituzione.

Il *mark-up* descrittivo è quello più interessante per l'*information retrieval*, e si identifica quasi perfettamente con il *mark-up* generico definito in precedenza. In questo caso il *mark-up* non dice al *software* cosa fare, ma descrive l'oggetto testuale nelle sue parti. Perché tale sistema funzioni, è necessario che il programma visualizzatore abbia prima ricevuto una serie di regole grafiche da applicare, per cui, per esempio, di fronte ad una citazione rientrerà il paragrafo e metterà il testo in corsivo. Il linguaggio SGML segue fondamentalmente questa filosofia: l'insieme di regole che il *software* di visualizzazione deve ricevere si definisce DTD (*document type definition*). Torneremo sul *mark-up* descrittivo più avanti, ma prima concludiamo la nostra breve rassegna.

L'ultima categoria di *mark-up* da considerare è il cosiddetto metamark-up: un sistema che fornisce all'utente la possibilità di interagire con il *mark-up* stesso, sul controllo dell'interpretazione da parte del *software*, sulla possibilità di estendere il vocabolario a disposizione della macchina. Più che di un vero e proprio sistema di *mark-up*, si tratta di una possibile applicazione che altri *mark-up* forniscono: in particolare quello procedurale, nei limiti della possibilità dell'elaboratore, e quello descrittivo.

Ma come si entra in 'interazione' con il *mark-up*? Probabilmente la maggior parte degli utenti di *word processor* utilizza continuamente questi sistemi senza neanche rendersene conto. Ciò è reso possibile dal fatto che il *mark-up* può rendersi invisibile ed essere completamente gestito dalla macchina. È il caso dei *word processor* più diffusi che utilizzano un sistema WYSIWYG. In questo caso, l'utente esegue delle operazioni direttamente sulla versione grafica del testo, che generano una risposta procedurale da parte del *software*, il quale, tuttavia, nasconde completamente la struttura interna del documento. La maggior parte dei programmi di scrittura (WordPerfect, Microsoft Word) sono in grado di mostrare questa struttura, ma limitano fortemente gli interventi sul *mark-up*, che in ogni caso si disinteressa dei contenuti; nel caso di Microsoft Word, poi, è impossibile

accedere ai codici del programma. Un caso evidente di *mark-up* visibile è, al contrario, quello delle pagine HTML, che permettono la visualizzazione della sorgente pagina che le ha generate.

A questo punto possiamo giungere ad alcune conclusioni interessanti per *l'information retrieval*. Abbiamo già affermato che il sistema che promette di più è quello descrittivo. In questo modo si descrive la natura di un testo, disinteressandosi della sua esposizione finale. L'SGML, in particolare, rappresenta il documento come una struttura ad albero, costituito da elementi che intrattengono relazioni gerarchiche e strutturali¹⁴⁸. Il vantaggio più evidente è quello della portabilità del testo, e della sua conservazione. Se sarà necessario modificare la forma grafica di un testo, perché sono subentrati nuovi standard stilistici, per la necessità di trasportare i testi su macchine più potenti (il problema della migrazione, che tratteremo in seguito), o semplicemente perché tale è la decisione dell'autore, non sarà obbligatorio riscrivere completamente il testo o il *mark-up*. Ciò che sarà necessario sarà una riscrittura delle regole di visualizzazione del testo. Come sostengono Coombs, Renear e DeRose,

With descriptive *mark-up*, properly tagged source files never require modification, and there is no such thing as *mark-up* maintenance.¹⁴⁹

Inoltre non si riscontrano, con questo sistema, i problemi di compatibilità del *software*: oggi, infatti, qualsiasi macchina, qualsiasi sistema operativo, qualsiasi programma di scrittura è in grado visualizzare un testo formattato in HTML. Quello della portabilità del testo non è un problema di secondo piano: una decina di anni fa Meyrowitz¹⁵⁰ sosteneva che le possibilità che l'ipertesto avrebbe offerto allo sviluppo culturale sarebbero state tali solo quando i prodotti ipertestuali fossero stati perfettamente compatibili in qualsiasi ambiente di lavoro. Oggi questo problema è stato risolto, sul *web*, proprio grazie alla diffusione dell'HTML, mentre persiste nei testi su CD-ROM che si oppongono ad una standardizzazione utilizzando sistemi di

¹⁴⁸ Cfr. **Fabio Ciotti**, "Testo rappresentazione e computer. Contributi per una teoria della codificazione informatica dei testi", in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p. 224.

¹⁴⁹ **James H. Coombs, Allen H. Renear, Steven J. DeRose**, op.cit., p.938.

¹⁵⁰ **Norman Meyrowitz**, "The missing link: Why We're All Doing Hypertext Wrong", in Edward Barrett (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press, 1989, p.107.

codifica proprietari. Ma i vantaggi più grossi si riscontrano nell'attività di *information retrieval*: in questo modo il *software*, infatti, è in grado di riconoscere la struttura di un documento, e, seppure parzialmente, la forma dei suoi contenuti. Si pensi ad un ricercatore cui interessano tutte le citazioni fatte dall'autore di questa ricerca: se questo testo fosse elaborato attraverso un *mark-up* descrittivo¹⁵¹, il sistema di ricerca potrebbe rintracciare, attraverso il *mark-up*, immediatamente, tutte le citazioni, tutte le frasi in neretto e quelle in corsivo.

Ripensando al modello cognitivo di Tonfoni cui si è fatto riferimento nel capitolo due, diviene possibile, per un autore, fornire un *mark-up* molto più ricco, evidenziando per esempio le frasi in cui sta spiegando qualcosa, quelle in cui sta narrando, quelle in cui sta ricapitolando. Si tratta, in ogni caso, di informazioni preziose per un sistema di *information retrieval*: un lettore potrebbe così evidenziare e confrontare particolare aspetti, e sviscerare la struttura testuale al di là della sua semplice manifestazione lineare e visiva. Citando ancora una volta gli autori del saggio sopra menzionato, questo sistema favorirebbe lo scrittore stesso, che potrebbe meglio rappresentare il suo processo di scrittura, identificare la struttura sottostante al suo elaborato, rendersi conto del proprio processo realizzativo:

One of the more subtle advantages of descriptive markup is it supports authors in focusing on the structure and the content of documents. Both presentational and procedural markup tend to focus authors' attention on physical presentation (...) Unlike descriptive *mark-up*, presentational and procedural *mark-up* fail to support the writer in developing the structure; even worse, they distract from the content.¹⁵²

La domanda che a questo punto sarebbe doveroso porsi è : se questo sistema incontra, pienamente, le possibilità tecnologiche e le prospettive cognitive, cosa ne ha impedito, sinora, la diffusione? Se davvero questo *mark-up* permette di ricercare strutture semantiche e favorisce il processo

¹⁵¹ Attualmente sto usando Microsoft Word, che utilizza un sistema di *mark-up* tipicamente procedurale e nascosto. La scelta è dovuta alla rigidità imposta alla versione grafica di una tesi di laurea. Non è da escludersi l'eventualità futura di prepararne una versione più 'trasportabile' e 'conservabile'.

¹⁵² James H. Coombs, Allen H. Renear, Steven J. DeRose, "Markup Systems and the Future of Scholarly Text processing", in *Communications of the ACM*, V.30, n°11, Novembre 1987, p. 943.

stesso di scrittura, tanto da essere definito anche *mark-up* contenutistico (*content mark-up*) come mai i principali *word processor* attualmente utilizzano sistemi di *mark-up* nascosti all'utente?

Per quanto riguarda i *word processor* di maggiore diffusione, essi hanno volutamente 'nascosto', come dicevamo prima, il *mark-up*, assecondando, se vogliamo, 'la pigrizia' dello scrittore attuale, il quale non vuole avere a che fare con problemi di *editing* (e questo è comprensibile) e spesso non è neanche in grado di fornire un'annotazione contenutistica ai suoi elaborati. Finché non sarà diffusa l'idea che il processo di scrittura è pianificabile, e richiede pertanto una certa capacità di autorappresentazione e progettazione, nonché la consapevolezza della mutevole struttura testuale, gli scrittori che utilizzano *word processor* continueranno a scrivere senza sfruttare a pieno le capacità della macchina e, quel che è più grave, le loro stesse capacità.

Per quel che concerne invece i sistemi di *mark-up* descrittivi, essi hanno effettivamente dei limiti: innanzitutto, la necessità di trovare un accordo sulle regole. La battaglia per gli standard è da sempre una delle più difficili da combattere in ambito tecnologico. Un sistema come SGML, per esempio, si limita a fornire delle regole di sintassi per il funzionamento dei visualizzatori; le annotazioni che ne costituiscono l'ossatura, le 'parole' insomma di questo linguaggio, vanno stabilite attraverso i DTD: se essi sono standardizzati, i sistemi sono compatibili, le macchine 'comprendono' quello che leggono. Viceversa, in mancanza di uno standard ratificato, il sistema si complica enormemente: se ogni autore inserisce nuove annotazioni, come, tanto per riportare l'esempio di Ressler¹⁵³, le annotazioni <ingredienti> <temperatura> per alcuni testi di ricette, in breve la capacità dei programmi di visualizzare ogni documento diventa insostenibile, generando una Babele elettronica in cui i computer conoscono la sintassi ma non le unità semantiche. Anziché facilitare la ricerca dell'informazione, questo sistema la complicherebbe fino a renderla impossibile. Esistono dei DTD standardizzati, come l'Electronic Manuscript Standard of the Association of American Publishers, ma essi non trovano tutti d'accordo. Si potrebbe infatti obiettare che degli standard troppo rigidi limitano le possibilità espressive di un autore: chi ha l'autorità per definire volta per volta le annotazioni da affiancare al testo? Queste posizioni differenti hanno fatto sì che purtroppo l'accordo sulle regole non sia stato trovato, almeno per quanto riguarda

¹⁵³ Sandy Ressler, *The Art of Internet Publishing, The Internet and Beyond*, Upper Saddle River, Prentice Hall, 1997, p. 170.

l'HTML. La sua prima versione standard è risultata infatti piuttosto carente. I due principali *browser*, allora, hanno intrapreso una gara nell'introdurre nuovi *tag*, nuove annotazioni possibili, che venivano incrementati in ogni nuova *release* dei prodotti, e che il comitato internazionale degli standard si limitava a ratificare con un certo ritardo. Come se ciò non fosse bastato, i *tag* introdotti hanno tradito la filosofia 'semanticista' dell'SGML, perché si sono limitati a fornire descrizioni sulle modalità espressive di un testo, trascurando completamente i contenuti. Ma anche gli utenti della rete, e in particolare i realizzatori di siti *web*, hanno le loro responsabilità a riguardo: spesso utilizzano l'HTML per illustrare una presentazione piuttosto che per descrivere un contenuto. È stato infatti dimostrato che alcuni *tag* di valore semantico elevato, come oppure , che dovrebbero annotare un particolare rilievo dato ad una proposizione oppure un'enfasi attribuita ad una parola o ad una frase, vengono utilizzati spesso per ottenere caratteri marcati o in corsivo piuttosto che per il loro contenuto semantico; così come si abusa dei tag di titolazione <H5>, <H6>, che sono usati per ottenere caratteri più piccoli o più grandi, anziché individuare una titolazione¹⁵⁴.

Per comprendere meglio queste problematiche, è necessario sottolineare che dietro ad un sistema di codifica c'è, come è inevitabile che fosse, una teoria cognitiva precisa¹⁵⁵. L'SGML, infatti, si è diffuso accompagnato e sostenuto dalla cosiddetta teoria gerarchica, per cui un testo è una gerarchia ordinata di oggetti di contenuto: capitoli, paragrafi, citazioni, enfasi, eccetera. Questi oggetti sono disposti in maniera gerarchica perché alcuni ne contengono altri, e perché, tra oggetti posti sul medesimo livello gerarchico, esiste una relazione lineare. Ora, è facile comprendere come una tale teoria faccia ampiamente affidamento sul concetto di genere, dal momento che è il genere testuale a determinare gli oggetti di contenuto testuale. E non sorprendono altresì le critiche che sono state rivolte a questo modello, che "nel riprendere le elaborazioni di natura funzionale della letteratura tecnica le trasforma in un asserto essenzialista circa la natura del testo"¹⁵⁶. È vero che il testo gerarchicamente strutturato è facilmente trattabile dagli strumenti informatici; ma non è vero che il testo è sempre gerarchicamente strutturato

¹⁵⁴ **Anna Stefani**, "Ci si può fidare della struttura HTML dei documenti Web?", in *Web Marketing Tools*, n°12, dicembre 1998, p.47.

¹⁵⁵ Si confronti, per questa sezione, **Fabio Ciotti**, "Testo rappresentazione e computer. Contributi per una teoria della codificazione informatica dei testi", in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997.

¹⁵⁶ **Fabio Ciotti**, op. cit., p. 233.

in maniera inequivocabile. I problemi relativi alla nozione di genere hanno dimostrato come ci siano testi che sfuggano ad una definizione di genere precisa; d'altra parte, tra gli studiosi non c'è neanche accordo su cosa sia un genere testuale, e su quanti ne esistano. Inoltre, le strutture relazionali tra gli elementi di un testo spesso si sovrappongono, perché lo stesso elemento può appartenere a diverse gerarchie. Da queste critiche si è sviluppato un modello teorico "pluralista", che rifiuta gli assunti "essenzialisti" della teoria gerarchica, per cui ogni testo è dotato di una struttura essenziale. Questo modello non abbandona l'idea di struttura gerarchica, ma accetta la possibilità che due elementi si sovrappongano in presenza di diverse prospettive analitiche. Si recupera insomma la possibilità di interpretazioni differenti, recuperando la nozione linguistica di livello testuale. Pertanto, le versioni aggiornate della TEI¹⁵⁷ prevedono la possibilità di contemplare elementi generici che possono occorrere in qualsiasi tipo di testo; elementi specifici di un determinato genere testuale; elementi addizionali prodotti da pratiche specializzate e prospettive analitiche particolari. Insomma, ogni studioso può costruire la sua 'vista' del testo modificandone il DTD.

Anche a questo modello non sono mancate critiche, relative soprattutto all'eccessiva informalità con cui si definivano le prospettive analitiche e la sovrapposizione di gerarchie. Pertanto si è giunti ad una terza revisione della teoria gerarchica che, abbandonando completamente ogni assunto essenzialista, prende in considerazione il concetto di sottoprospectiva: se due oggetti testuali evidenziati da una prospettiva teorica si sovrappongono, ciò significa che essi appartengono a due sottoprospective diverse. Il testo diventa un sistema a più livelli, ognuno dei quali corrisponde ad un punto di vista dell'osservatore. Anche le sottoprospective danno luogo a gerarchie. Dal "platonismo" esistenziale delle prime teorie, la ricerca sembra muoversi verso prospettive sempre più costruttiviste e relativiste, per cui la forma strutturale perderebbe valore ontologico, dal momento che ogni autore produce un'istanza del testo diversa. Prendendo in considerazione questo punto di vista estremo, tuttavia, si perde completamente la possibilità di trattamento computazionale del testo.

La questione che si nasconde dietro un velo di informatizzazione ed elaborazione elettronica è di portata ben più ampia, e risale ai fondamenti stessi della filosofia del linguaggio. Molto sinteticamente possiamo osservare

¹⁵⁷ Text Encoding Initiative: uno schema di codifica basato sulla sintassi SGML rivolto specificamente al dominio umanistico.

che più il modello descrittivo del testo si raffina, più complessa diventa la sua trattazione informatica, sino a diventare irrealizzabile. Occorre dunque ricercare una soddisfacente via mediana, essendo ben consapevoli delle rinunce che qualsiasi scelta comporta. È senz'altro vero che le interpretazioni possibili di un testo sono ben più numerose di quelle progettate dall'autore. Ogni codifica informatica permette di costruire modelli formali di un testo che ne descrivono alcune caratteristiche a un determinato livello, ma non può mai coincidere con il testo stesso nella sua pienezza. Scrive Ciotti:

Una rappresentazione codificata di un testo è vera se è internamente coerente, accettabile razionalmente nell'ambito di una teoria, in grado di rappresentare i fenomeni testuali rilevanti nel contesto di quella teoria o prospettiva metodologica¹⁵⁸.

Limitatamente a quanto concerne la manualistica tecnica, la codifica informatica elaborata dall'autore, o dal curatore del testo, e quindi il modello interpretativo soggiacente, per quanto contestabile sia, fornisce pur sempre delle informazioni valide per il reperimento dell'informazione. Se poi il lettore si rifiuta di utilizzare queste informazioni, è liberissimo di percorrere delle strade alternative: questo non ci autorizza a buttare al vento tutte le possibilità offerte dalla codifica informatica.

Un esempio di queste possibilità è mostrato egregiamente dal saggio di Colombo¹⁵⁹. Secondo l'autrice la codifica informatica è un atto di traduzione e commento testuale, che deve tenere in considerazione la presenza di elementi di significazione non discreti, ma continui, caratterizzanti il piano contenutistico del testo. Il piano espressivo, e quindi tipografico, che è invece di natura discreta, non sempre riesce a disambiguare questi elementi: la presenza del corsivo, in un testo, può identificare una parola straniera, oppure un titolo. Riprendendo la nozione di testo come entità complessa stratificata, Colombo sottolinea come i fattori di significazione siano compresenti, si sovrappongono e si intersechino. Mentre in un testo scritto un lettore esperto può cogliere questi diversi elementi testuali

¹⁵⁸ **Fabio Ciotti**, "Testo rappresentazione e computer. Contributi per una teoria della codificazione informatica dei testi", in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p. 243.

¹⁵⁹ **Chiara Colombo**, "Teoria e prassi della codifica del testo nell'analisi computerizzata dell'epistolario di J.G. Hamann", in Patrizia Bellman Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997.

contemporaneamente, la codifica informatica, che opera su basi discrete, necessita l'esplicita codifica di tutti questi dati, richiedendo pertanto che il testo venga "analizzato e scomposto in tutti gli elementi logici, grafici, dell'espressione che il ricercatore ritiene necessari e sufficienti a riprodurlo fedelmente in vista di future elaborazioni"¹⁶⁰. Per ciò che concerne l'aspetto fisico di un testo (presenza di capitoli, paragrafi, titoli) il lavoro di codifica è più semplice e oggettivamente indiscutibile; molto più complessa è la codifica dei valori funzionali che ogni parte del testo assume rispetto alle altre; in questo caso il giudizio, dell'autore o dell'analista, per quanto condivisibile e giustificabile, è comunque soggettivo, o inter-soggettivo. Ancora più impegnativa è la codifica semantica¹⁶¹ vera e propria, relativa al discorso del testo, al messaggio in quanto tale, chiamando in causa tutta la competenza linguistica personale del codificatore. Colombo descrive poi la codifica effettuata, date queste premesse, su un corpus di lettere di Hamann. Ciò che maggiormente ci preme rilevare è che questo tipo di codifica permette al lettore una serie di strumenti informatici di *information retrieval* di potenza rilevante: è possibile ricercare, confrontare e combinare dinamicamente tutti gli oggetti testuali in cui il testo è stato codificato, per tutti i livelli disponibili, da quello tipografico a quello morfosintattico, sino a quello semantico. Se proprio vogliamo trovare un limite al lavoro presentato da Colombo, esso sta nella scelta di un sistema di codifica poco trasportabile e molto specialistico come i formati GIRSCE e DBT.

In precedenza abbiamo già accennato, trattando della metodologia di Tonfoni CPP-TRS, di alcune sue possibili applicazioni informatiche, relative in particolare alla realizzazione di sistemi di codifica. L'impianto teorico della metodologia è infatti alla base di un linguaggio di *mark-up* da essa derivato, il CTML¹⁶², pensato per l'annotazione e il collegamento dei documenti sul *web*.

Il linguaggio è stato progettato con il fine di migliorare la visibilità del

¹⁶⁰ Chiara Colombo, op.cit., p. 302.

¹⁶¹ L'autrice definisce questo livello semantico "piano dell'espressione", riferendosi a ciò che in questo lavoro viene invece definito piano del contenuto. Chi voglia confrontare i testi tenga presente questa non irrilevante differenza, dovuta evidentemente alle diverse scuole linguistiche cui si fa riferimento.

¹⁶² Context Transport Mark-up Language. Si veda a riguardo **Graziella Tonfoni**, "Intelligent control and monitoring of strategic documentation: a complex system for knowledge miniaturization and text iconization", in *The Proceedings of the ISIC/CIRA/ISAS 98 Conference*, Gaithersburg, National Institute of Standards and Technology, U.S. Dept. of Commerce, MD, 1998, e, della stessa autrice, "On augmenting documentation reliability through communicative context transport" in *The Proceedings of the 1999 Symposium on Document Image Understanding Technology*, Annapolis, MD, 1999.

contenuto testuale e del contesto, aspetti, come abbiamo visto, purtroppo frequentemente trascurati. Si basa su una serie di schemi visivi e una visualizzazione dinamica, che possono essere combinate per rappresentare il tipo di informazione contenuto in ogni paragrafo. La scelta delle categorie a partire dalle quali classificare il contenuto ovviamente è stata a lungo testata e non può mai dirsi definitivamente conclusa, perché essa va adeguata ad un numero continuamente cangiante di variabili quali le necessità dello scrittore, il periodo storico, il tipo di messaggio veicolato dal testo, eccetera. L'utilizzo di marche semantiche trattabili in maniera automatizzata si prefigge, come primo obiettivo, la facilitazione delle attività di reperimento dell'informazione. Le marche contestuali visualizzabili dal CTML sono le seguenti: 'funzione comunicativa o tipo', designata da dieci diversi simboli, ognuno dei quali rappresenta una diversa funzione; 'intenzione comunicativa o stile', designata da undici simboli; 'avvicendamento comunicativo dei turni o delle azioni da prendere sul documento' (quattro simboli). Questi simboli forniscono agli utenti alcuni indizi interpretativi, che altrimenti sarebbero invisibili, favorendo la comprensione degli stessi.

Ci avviamo alla conclusione di questa sezione trattando un problema di diversa natura che è stato già, più o meno trasversalmente, annunciato: quello della conservazione dell'informazione. È un problema fondamentale, come è ovvio, per l'*information retrieval*, di cui mina le stesse fondamenta, anche se nel campo della ricerca tecnologico non pare sia stato ancora degnamente affrontato.

3.6. La 'migrazione' dell'informazione

Un libro stampato su carta comune comincia a deteriorarsi dopo un'ottantina di anni, e i bibliofili conoscono bene le difficoltà che comporta conservare in buono stato un libro secolare. Nonostante le ottimistiche affermazioni che accompagnano il lancio di nuovi supporti elettronici, non ultime quelle sul DVD-ROM¹⁶³, un'evoluzione del CD-ROM in grado di memorizzare più di 4 Gb di dati, la situazione nel campo della conservazione dei supporti elettronici è molto più grave: l'informazione può essere infatti conservata senza perdite, ma non senza costi, come spiega Lesk¹⁶⁴. L'autore infatti evidenzia due generi di problemi relativi alla conservazione

¹⁶³ Digital Video Disk Read Only Memory.

¹⁶⁴ Michael Lesk, *Preserving Digital Objects: Recurrent Needs and Challenges*, <http://community.bellcore.com/lesk/auspres/aus.html>. Si vedano anche gli Atti del Convegno internazionale "The Future of Memory", tenuto presso l'Università di San Marino il 21-23 maggio 1999.

dell'informazione su supporti elettronici: la deperibilità dei supporti stessi e, la rapida obsolescenza dell'apparato *hardware-software* in grado di gestirla.

Il primo problema non è assolutamente nuovo, perché anche il supporto cartaceo è destinato a subire processi di invecchiamento che lo rendono, a lungo andare, inutilizzabile. L'età media di un supporto elettronico è molto variabile, e tuttavia quasi mai permette di migliorare le 'prestazioni' della carta tradizionale. Se un CD-ROM, infatti, si presenta come un supporto piuttosto duraturo, dalla durata media di circa cento anni, un supporto magneto-ottico non ha più di trent'anni di vita, un nastro a 8 millimetri difficilmente supera i dieci anni, e un nastro magnetico a 9 tracce, piuttosto diffuso negli anni sessanta come supporto di memorizzazione, non ha più di uno o due anni di vita. Difatti, i nastri prodotti negli anni sessanta non sono più utilizzabili. Tuttavia, quello della deperibilità dei supporti elettronici dell'informazione non è il più serio dei problemi che chi catalogherà informazioni in futuro dovrà affrontare. Infatti il problema della deperibilità dell'*hardware* e del *software* di lettura è ben più grave.

A patto che i CD-ROM prodotti attualmente siano ancora leggibili fra cento anni, chi ci garantisce che per quella data saranno ancora disponibili lettori e *software* in grado di gestirli? La questione è legata alla complessità generata dalla funzionalità elettronica digitale. L'esempio che Desk fa è eloquente: se si rompe una meridiana, con un compasso e un po' di pazienza chiunque è in grado di ripararla. Se invece si tratta di un orologio digitale, pochissimi utenti saranno in grado di manipolarlo. La tecnologia digitale frappone un 'velo' fra il lettore e l'informazione, costituito appunto dal *software*. Questo 'velo' non solo richiede competenze specifiche per essere gestito, dal momento che utilizzare un CD-ROM con un computer è obiettivamente più complesso che sfogliare un libro; rischia anche di interrompere del tutto il contatto tra utente e informazione. Si consideri, per esempio, l'informazione memorizzata negli anni sessanta: abbiamo detto che molto probabilmente i supporti saranno ormai deteriorati. Ma anche se qualcuno si fosse conservato, sarebbe inutilizzabile, perché richiederebbe l'utilizzo di macchine che non esistono più e di sistemi operativi e di linguaggi ormai caduti in disuso. La soluzione allora potrebbe essere quella di conservare le macchine. A parte il fatto che ciò sarebbe poco economico, perché ogni biblioteca dovrebbe fornirsi di tutti i tipi di computer prodotti dal 1960 circa ad oggi, ciò si rivelerebbe comunque non fattibile, perché oltre agli elaboratori, occorrono tecnici in grado di farli funzionare, e oggi

pochissimi sarebbero in grado di trovare un informatico che sappia far funzionare un IBM 7090. Abbiamo visto che l'*information design* ci permette di considerare la documentazione tecnica un flusso continuo, dal momento che sono necessari continui aggiornamenti. Tuttavia, non è davvero proponibile il mantenimento di computer e *software* obsoleti per leggere vecchi documenti: la manualistica necessaria al funzionamento dei computer finirebbe per occupare parte consistente della biblioteca.

Le soluzioni proponibili sarebbero due: il ritorno alla cara e vecchia carta, oppure la cosiddetta 'migrazione' dell'informazione. Per quanto riguarda la prima soluzione, essa va scartata immediatamente se pensiamo che la maggior parte della produzione testuale attuale non raggiunge mai la fase di stampa: si pensi ai milioni di pagine di *e-mail*, *news-group* e *homepage* che Internet produce continuamente. Il quantitativo necessario per immagazzinare una tale mole di informazione sarebbe inimmaginabile. Sempre che ne valga la pena, si potrebbe obiettare: non possiamo pretendere di conservare tutto, sommergendo i nostri eredi di informazione inutile. Un processo di selezione dell'informazione che merita di essere conservata, si potrebbe aggiungere, va attuato comunque; esso si profila indispensabile per la letteratura, cioè per testi di un certo livello, figuriamoci se non dovrebbe essere necessario per la testualità più 'povera'. Riferendosi all'importanza di mantenere un canone della letteratura occidentale, il critico Bloom scrive:

Who reads must choose, since there is literally not enough time to read everything, even if one does nothing but read. Mallarmé's grand line -"the flesh is sad, alas, and I have read all the books" has become a hyperbole.¹⁶⁵

Condividiamo qui la riflessione di Desk, che ricorda che uno dei pericoli del mondo digitale è la credenza di poter conservare tutto, senza renderci conto dei costi che comporta la catalogazione del materiale, necessaria affinché qualcuno la recuperi. Anche Cazalè Bèrard e Mordenti riflettono su questo punto:

È possibile prevedere e preventivare la memorizzazione elettronica completa di una data

¹⁶⁵ **Harold Bloom**, *The Western Canon: the Books and School of the Ages*, New York, Harcourt Brace, p.15.

tradizione testuale, di un corpus di opere, di un ambito di sapere attraverso i documenti della sua esistenza e del suo sviluppo? È utile e opportuno "conservare" (ri-codificando, cioè editando informaticamente) senza scelta, senza discriminare, senza programmazione scientifica?¹⁶⁶

E ancora, poco più avanti:

(...)Non si dà "conservazione" dotata di valore euristico che non comporti anche una "selezione" degli oggetti di studio basata su criteri definiti dalla comunità scientifica degli specialisti.¹⁶⁷

Non approfondiamo tuttavia la questione, perché in ogni caso, ci saranno molte buone ragioni portate a difesa di chi vuole conservare tutto, così come ce ne saranno di ottime per chi vuole operare una cernita. Dal punto di vista dell'*information retrieval*, comunque, vale la pena sottolineare che un conto è selezionare l'informazione che si reputa più rilevante per i posteri, un conto è rendere impossibile l'accesso a certa informazione per motivi tecnologici. L'opzione della conservazione su carta va messa da parte comunque, e definitivamente, per una semplice ragione: a patto che si decidesse di operare una selezione di ciò che va conservato, ci sarebbero comunque dei testi che non potrebbero essere conservati su carta: gli ipertesti multimediali, per esempio.

La soluzione dunque deve andare rinvenuta nella migrazione dell'informazione. Con questo termine Desk definisce il passaggio dell'informazione dal supporto fisico sulla quale è conservata ad un supporto più recente.

Libraries must anticipate that in a digital world, information must be copied regularly. The word "migration" is used to indicate that the copying may not be exact. It is contrasted with simple copying, as might be done to refresh magnetic tape

¹⁶⁶ **Cazalé Bérard Claude, Mordenti Raul**, "La costituzione del testo e la «comunità degli interpreti»", in Patrizia Bellman Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.28.

¹⁶⁷ **Cazalé Bérard Claude, Mordenti Raul**, op.cit., p.30.

copies on the identical kind of media and tape drives(...)Migration to newer technologies may involve some complexity.¹⁶⁸

La diffusione di standard di *mark-up* descrittivo, come esposto in precedenza, può ridurre di molto la difficoltà della migrazione, che, ben inteso, non è un processo semplice; richiede competenze informatiche e cognitive. Si pensi infatti ad un ipertesto multimediale: esso può contenere testo, per esempio in formato HTML, immagini nello standard di compressione JPEG¹⁶⁹, filmati video in formato MPEG¹⁷⁰-1, audio in formato CD o MP3¹⁷¹. Ebbene, da un punto di vista strettamente computazionale, sarebbe comodo 'migrare' queste categorie separatamente. Oppure, per risparmiare tempo, decidere di salvare solo il testo; oppure salvare tutto, perdendo però i codici che permettono l'interazione del materiale. È ovvio che queste soluzioni sono aberranti da un punto di vista linguistico. Se quello dell'ipertesto è un vero e proprio nuovo linguaggio, nato dalla confluenza e dall'integrazione di più codici, separarne i componenti equivarrebbe più o meno a copiare tutte le illustrazioni di un fumetto, e poi, a parte, i testi contenuti dai fumetti. Come Desk suggerisce, l'unica vera soluzione possibile è la cooperazione: cooperazione tra biblioteche, che si suddividano il compito di migrare alcuni testi, scambiandosi gli altri; cooperazione tra gli apparati legislativi, che considerino le difficoltà di natura legale che la copia di un testo da un formato ad un altro può comportare, ma anche i costi notevoli, che solo un'entità statale è in grado di supportare; cooperazione tra autori e informatici, nella consapevolezza che il lavoro dei primi rischia di essere vanificato senza l'ausilio dei secondi; cooperazione tra programmatori e produttori di *software*, perché si raggiungano degli standard comuni efficaci. Dobbiamo insomma cominciare a preoccuparci adesso di rendere realizzabile nella maniera più semplice possibile le attività di *information retrieval* di domani.

3.7. Brevi cenni sulle applicazioni dei sistemi esperti per l'information retrieval

A conclusione di questo capitolo sugli strumenti informatici per il reperimento dell'informazione, è opportuno fare almeno un accenno ad una

¹⁶⁸ Michael Desk, op.cit., p. 3.

¹⁶⁹ Joint Photographic Experts Group.

¹⁷⁰ Motion Pictures Experts Group.

¹⁷¹ MPEG-1 Layer III: un formato di compressione audio che nasce come variante dell'MPEG audio.

serie di studi che si pongono come obiettivo l'utilizzo di elementi di Intelligenza Artificiale per la creazione di ipertesti o database 'esperti', in grado cioè di fornire un aiuto 'intelligente' all'utente sulla base di una serie di funzioni per le quali sono stati predisposti. Si tratta quasi sempre di prototipi che difficilmente escono dai laboratori di ricerca nei quali sono stati sperimentati per raggiungere la grande diffusione; almeno per il momento. Patricia Ann Carlson si occupa appunto di interfacce intelligenti implementate sulla base di meccanismi di "*idea processing*" programmabili. Il saggio¹⁷² si occupa della versione elettronica di alcuni manuali di istruzioni, i quali, per essere più efficaci dei loro equivalenti cartacei, devono offrire miglioramenti, quali una flessibilità maggiore, una riduzione degli spazi di immagazzinamento, un sistema di sviluppo e mantenimento dei documenti comodo. Per Carlson, l'ipertesto è più efficiente per il recupero dell'informazione rispetto al *database* tradizionale perché ricorda più da vicino la rappresentazione della conoscenza dell'Intelligenza Artificiale, con le sue reti semantiche, i *frame* procedurali, eccetera. Le applicazioni 'intelligenti' all'ipertesto suggerite si dividono in tre categorie.

La prima categoria è quella cronologicamente venuta prima, basata sul metodo delle *query* (richieste): un miglioramento in questa direzione è offerto dall'uso di ricerche di parole chiave "appesantite, rinforzate"¹⁷³ (*weighted key-word search*). Il sistema non risponde ad una *query* fornendo una serie di occorrenze testuali, ma una serie di nodi, alcuni punti di accesso all'ipertesto che rimandano a concetti correlati: ciò evita la falsa impressione di completezza che può generare una ricerca *full-text*. A dieci anni di distanza dal saggio, che è dell'89, ci pare di poter dire che questo sistema è ampiamente utilizzato in rete, mentre per i CD-ROM si continua a preferire la ricerca tradizionale. Un'altra miglioria apportabile è quella di fornire sistemi dotati di menù di scelta, che riducono il sovraccarico cognitivo dell'utente, inseriti direttamente nel testo, per non smarrire la contestualità. L'utente può così volta per volta decidere in quale direzione muovere la ricerca; tale sistema, ottimo sulla carta, è però praticamente irrealizzabile, come ammette Carlson, per l'enorme lavoro che richiede all'autore.

Se i sistemi di ricerca della prima categoria aiutano l'utente a selezionare

¹⁷² Patricia Ann Carlson, "Hypertext and Intelligent Interface for Text Retrieval", in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press, 1989.

¹⁷³ La traduzione letteraria è piuttosto inefficace: il termine *weighted* infatti indica la condizione di un oggetto che è stato sistemato in maniera tale da favorirne l'utilizzo da parte di qualcuno.

un punto di accesso al testo, quelli appartenenti alla seconda si sforzano di aiutarlo a visualizzare graficamente la struttura dell'archivio, fornendo così un'alternativa agli elementi paratestuali che, come si è già detto, scarseggiano per la testualità elettronica. Non mancano esempi di sistema di ricerca riferiti a questa categoria: la cosiddetta visuale ad occhio di pesce, per cui i termini di maggiore rilevanza sono posti al centro della visuale, con dimensioni maggiori; quelli periferici sono presentati con un formato più piccolo ai margini dell'immagine; i filtri grafici, che presentano l'informazione sotto forma di relazioni grafiche; i *browser* grafici, l'uso cioè di immagini gerarchiche di cartelle che ne contengono altre¹⁷⁴. I sistemi di ricerca della seconda categoria partono da metafore del mondo reale, sono pertanto piuttosto immediati nell'utilizzo, favoriscono l'intuizione riducendo le inibizioni che alcune interfacce complesse possono generare in utenti meno esperti.

I sistemi della terza categoria sono quelli più avanzati che attingono maggiormente dalle conoscenze degli studi sull'Intelligenza Artificiale. Si tratta di sistemi esperti che non utilizzano gli algoritmi di ricerca convenzionali, ma "elaborazioni di alto livello, euristiche, simboliche"¹⁷⁵. Costituiti da un *database* e da un sistema di recupero dell'informazione, questi programmi applicano a quest'ultimo alcuni componenti 'intelligenti'. Il sistema esperto è fornito infatti di una serie di informazioni conoscenze, elementi di "heuristic expertise" che gli permettono di realizzare delle inferenze. Questo sistema logico di inferenze può essere posto a monte del processo di ricerca dell'informazione, e quindi selezionare subito la parte di *database* che contiene la documentazione più interessante per l'utente; può essere indistinguibile dall'ipertesto, con cui forma un tutt'uno, costruendo automaticamente dei collegamenti ipertestuali tra i documenti di cui dispone; può essere infine distribuito, costituito cioè da tanti piccoli sistemi esperti, ognuno costruito con regole di base diverse e richiamabili dall'utente in qualsiasi punto dell'esplorazione testuale.

Alcune di queste prospettive potrebbero apparire ai più scettici come appartenenti più a progetti fantascientifici che ad applicazioni reali. In effetti, questo tipo di *software* 'intelligenti' richiede un complesso lavoro di programmazione, degli alti costi quindi, che non possono essere

¹⁷⁴ La maggior parte dei personal computer attuali utilizzano sistemi simili per la gestione dei *file*.

¹⁷⁵ **Patricia Ann Carlson**, "Hypertext and Intelligent Interface for Text Retrieval", in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press, 1989, p.69.

ammortizzati con una produzione in massa perché spesso i sistemi esperti funzionano solo nel dominio per i quali sono stati progettati, poco adatti quindi ad una diffusione generale. La difficoltà nella realizzazione di questi sistemi dipende dall'enorme dispendio di energie per i progettisti richiesto dalla programmazione di una macchina in grado di effettuare inferenze logiche valide. Pertanto, i supporti per la conservazione dell'informazione pensati per utilizzo di massa, raramente fanno uso di tali strumenti, e non sono stati presi in considerazione neanche nell'esperienza che verrà analizzata nella seconda parte di questo lavoro. Non è qui tuttavia possibile addentrarci nelle questioni relative all'Intelligenza Artificiale, una materia 'giovane' ma con notevoli risultati alle spalle; ci è sembrato opportuno farne comunque un cenno, perché i sistemi esperti rappresentano uno dei possibili orizzonti futuri per l'*information retrieval*.

4. Dalla teoria alla realizzazione pratica: studio di un caso

“Ho fatto un corso di lettura veloce, ho imparato a leggere a piombo, trasversalmente la pagina, e ho potuto leggere *Guerra e Pace* in venti minuti. Parlava della Russia”
(W.Allen)

4.1. La genesi del testo compresso

Dopo aver fornito un quadro complessivo sulle prerogative linguistiche, semiotiche ed informatiche dell'*Information Retrieval*, passiamo adesso all'analisi, già ampiamente preannunciata, di un lavoro editoriale nella cui produzione si è cercato di realizzare quanto teorizzato dalla ricerca più avanzata nel settore, e in cui il sottoscritto, in qualità di curatore, ha avuto modo di osservare da vicino le fasi di realizzazione.

Cominciamo con il descrivere il punto di partenza: si era in possesso di quattro volumi già pubblicati di Graziella Tonfoni, ma fuori dal commercio e quindi non disponibili da parte dei lettori. I quattro testi esplorano le possibilità della metodologia CPP-TRS che abbiamo già avuto modo di citare, ma questo è irrilevante per gli scopi della nostra discussione: ai fini dell'intervento di *information design* che è stato operato, essi avrebbero potuto trattare anche altri tipi di contenuto, purché in modo sistematico e tecnico. Un discorso differente sarebbe valso nel caso di opere letterarie: in un romanzo o in un saggio critico l'intervento dell'*information designer* è di natura diversa, e in questa sede non sarà preso in considerazione. Da questi quattro volumi si voleva ottenere un unico testo assai più sintetico, in grado però di fornire strumenti efficaci per il reperimento dell'informazione. Occorreva perciò lavorare sui quattro testi originari, elaborarli, modificarli in qualche misura. È importante rilevare che i tre curatori non erano chiamati a riassumere i testi nella loro completezza, ma piuttosto ad evidenziarne alcune parti rilevanti secondo linee guida definite dall'autrice stessa. Alcune sezioni dei quattro volumi base, infatti, sono state messe da parte a priori, e si è stabilito che il nuovo testo compresso dovesse prendere in considerazione soltanto la parte dei quattro volumi riguardanti gli strumenti della metodologia definiti da Tonfoni *Tele Testuali*, *Oggetti Testuali*, *Macchine*

Testuali e Segnali e Simboli Testuali. L'attività è inizialmente partita come esercitazione didattica. Solo in un secondo momento, quando si è prospettata la possibilità di darne un seguito editoriale, si sono tenuti presenti da una parte le possibilità offerte dalla multimedialità, dall'uso degli strumenti informatici, e, dall'altra, i limiti imposti dalle regole economiche che sovrintendono la produzione editoriale. Un progetto didattico, nel momento in cui supera i confini dell'università per entrare nel mercato editoriale, deve prendere in considerazione una serie di costrizioni, limiti, ma anche possibilità che prima non si ponevano. Non si dimentichi infatti che

L'editore è un imprenditore a tutti gli effetti e, anche se qualcuno vorrebbe smentire, possiamo senza dubbio dire che lo scopo del suo operato è il profitto. (...) Fare l'editore significa, dunque, essere imprenditore e, come tale, compiere delle scelte strategiche riguardanti il prodotto, prima di tutto, quindi il mercato al quale indirizzarsi, la ricerca dei collaboratori, dei fornitori, la scelta tecnologica e la creazione di un'immagine aziendale (...)¹⁷⁶

Nel caso in questione, ampia libertà è stata lasciata a noi curatori per la realizzazione del prodotto finale e le scelte strategiche, ferme restando le prerogative di un editore riguardanti la produzione fisica, la scelta dei fornitori e dei distributori, il lancio pubblicitario, eccetera. Va rilevato inoltre un elemento di importanza non secondaria, e cioè che il testo si è posizionato in una collana già definita, con tutti i vantaggi e i limiti che ciò comporta¹⁷⁷, e ha potuto utilizzare un'immagine aziendale già consolidata.

¹⁷⁶ **Giuliana Barzon, Patrizia Bertini, Stefania Sordi**, *Trovar lavoro in editoria. Da freelance a imprenditore. Una guida completa alle professioni del libro*. Milano, Editrice Bibliografica, 1999, p.13.

¹⁷⁷ Una collana è un elemento paratestuale molto importante, perchè fornisce al lettore delle informazioni sul testo ancora prima che cominci la lettura, e quindi già durante la sua esposizione in libreria. Una collana rinomata è uno strumento per l'editore in quanto consente di diversificare il prodotto, ma anche una garanzia per il lettore, che può aver già letto e apprezzato altri testi della collana, oppure può stimare a tal punto il curatore della stessa da fidarsi completamente delle sue scelte. Un testo inserito in una collana è un testo che ha superato già un primo esame, una prima scrematura, da parte del direttore di collana, che garantisce con il suo nome, oltre a quello della casa editrice, sulla qualità del prodotto. È evidente però che l'inserimento in collana presenta anche alcuni limiti: i testi appartenenti ad una collana dovranno possedere delle caratteristiche comuni: contenutistiche, perchè il testo deve indirizzarsi al pubblico usuale della collana, ai suoi estimatori, e formali, di formato, perchè il testo spesso deve adattarsi al formato standard della collana, alle sue caratteristiche grafiche e tipografiche che ne garantiscono la riconoscibilità. Cfr. **Giuliana Barzon, Patrizia Bertini, Stefania Sordi**, op.cit., p.15.

Prima ancora di stabilire quale tipo di intervento operare sui testi base e quali supporti tecnologici utilizzare, si è fatto il punto sull'utente finale, sul pubblico di lettori, cui ci si voleva indirizzare, della collana in cui il libro sarebbe stato inserito. La collana, *I moduli* della Cuen, si rivolge ad un pubblico interessato alle nuove tecnologie, alle nuove possibilità offerte alla comunicazione, ad argomenti insomma talvolta anche piuttosto tecnici, ma mai specialistici ed esclusivamente pensati per addetti al settore. La collana infatti si presenta come una serie di testi tascabili, piuttosto economici dunque, proprio per raggiungere un bacino di utenza vasto. Perciò il nostro testo finale doveva essere un manuale, agile e di pratica consultazione, che potesse introdurre e guidare il lettore nella conoscenza progressiva della metodologia di scrittura multimediale ideata da Tonfoni. Il libro doveva rivolgersi ad un lettore che volesse conoscere in modo chiaro e veloce gli strumenti fondamentali e gli ambiti d'utilizzo della metodologia. La collana richiedeva, non è banale notarlo, che il prodotto finale fosse un libro; ma ciò non escludeva l'utilizzo di qualche supporto multimediale al libro che si affiancasse.

Ma procediamo con ordine: si è detto che eravamo in possesso di quattro lavori, di cui potevamo disporre liberamente perché i diritti d'autore¹⁷⁸ erano ritornati in possesso di Tonfoni che ci aveva garantito la sua disponibilità per l'utilizzo. Il nostro obiettivo era la realizzazione di un quinto volume, cartaceo, di dimensioni tali da poter essere inserito in una collana di tascabili, che contenesse gli elementi fondamentali dei primi quattro. Il compito era reso difficoltoso dalla natura dei quattro testi, i quali, essendo stati pensati per tipologie di pubblico diverse, presentavano degli elementi di differenziazione stilistica. Vediamo in breve queste caratteristiche¹⁷⁹. Il primo di essi, *Scrittura multimediale*, pubblicato nel 1991, si rivolge esplicitamente alla classe docente. Fornisce strumenti e suggerimenti volti all'utilizzo della metodologia CPP-TRS da parte degli insegnanti, con una rimarcata predominanza di elementi pedagogici e formativi. Da ciò ne consegue la costruzione del testo in maniera piuttosto ridondante, caratterizzata da un ritmo volutamente lento, al fine di permettere al lettore la comprensione di una serie di contenuti di base fondamentali. Il secondo, *La*

¹⁷⁸ Si fa ovviamente riferimento qui ai diritti di utilizzazione economica dei testi, ceduti dall'autrice all'editore, e non ai diritti morali dell'autrice, che, come previsto dal nostro ordinamento (legge sul diritto d'autore, n° 633 del 22.4.1941), sono inalienabili.

¹⁷⁹ Per questa parte, si sono utilizzati elementi e riflessioni già espressi nella prefazione e nell'introduzione al libro di **Graziella Tonfoni**, *Il design della scrittura multimediale*, Napoli, Cuen, Città della Scienza, 1999.

comunicazione aziendale come arte visiva, pubblicato anch'esso nel 1991, si rivolge invece principalmente ad un'utenza aziendale, con riferimenti alla applicazione della metodologia in attività esplicitamente imprenditoriali come il *problem-solving*, il *group working*, il *decision-making*. L'opera è stata pensata ed elaborata per una lettura ad alta voce, con le particolarità che tale scelta determina e che si riflette sensibilmente sullo stile espressivo. Il terzo volume, *Partitura, solfeggio, movimento*, pubblicato nel 1992, rappresenta una seconda tappa, rispetto ai primi due volumi, nel percorso esplicativo della metodologia CPP-TRS. Quest'opera presuppone la lettura di almeno una delle prime due, o di entrambe, in quanto costituisce la prosecuzione di alcune di idee portanti già esposte precedentemente. Dunque un testo inequivocabilmente dipendente da un altro, di cui si dispone come prosecuzione ed approfondimento; il quarto e ultimo volume, infine, *Abitare il Testo*, è stato completato per la stampa nel 1993 e pubblicato nel 1995. L'opera in questione rappresenta una sintesi dei tre libri che l'hanno preceduta, ed è stata pianificata e sviluppata dall'autrice in maniera non lineare, cioè ipertestuale. Si tratta dunque di un iperlibro, dotato di *link* interni, pensato anche in questo caso per una lettura ad alta voce. I quattro volumi si rivolgono pertanto a tipologie di pubblico diverse, utilizzano stili espressivi diversi, presentano delle ridondanze volute, delle ripetizioni progettate, degli accrescimenti: non si ritrovano delle discordanze interne, poiché si tratta di quattro volumi che rappresentano una metodologia coerentemente strutturata, ma alcuni 'aggiustamenti di tiro', inevitabili in un lavoro che si sviluppa nel tempo e può generare ripensamenti e perfezionamenti. Appariva evidente come i quattro volumi non fossero considerabili quattro 'puntate' della stessa opera, leggibili in sequenza. Inoltre essi presentavano una caratteristica che 'strideva' con il prodotto che invece noi volevamo realizzare. La caratteristica cui si fa riferimento è la natura dei quattro testi: una natura spiccatamente formativa, dal momento che i volumi erano stati costruiti per insegnare, mentre il nostro prodotto doveva essere prima di tutto informativo. Il nostro scopo era infatti mettere il lettore a conoscenza della metodologia, lasciandogli poi la libertà di approfondirne o meno lo studio. Bisognava trovare una soluzione che non eliminasse la possibilità di usufruire completamente dei quattro volumi, e al tempo stesso non fosse una semplice riproposizione degli stessi ma offrisse al lettore qualcosa di più. Da queste riflessioni si poteva escludere l'idea di fare del quinto volume un riassunto dei primi quattro: in questo modo, infatti, si sarebbe dovuto amalgamare lo stile, si sarebbe sicuramente guadagnato in

leggibilità e compattezza, ma sarebbero completamente andate perdute le differenze espressive e stilistiche dei quattro volumi, che rappresentavano un bagaglio di informazione non trascurabile. E poi la facilità di lettura non era un elemento indispensabile del progetto, dal momento che si voleva ottenere un manuale da consultare anche in maniera non lineare, e non un saggio esplicativo: il modello era insomma più quello di un manuale di istruzioni, che di una guida ragionata all'uso. Per cui l'ipotesi della sintesi è stata scartata, sostituita da quella dell'estrazione testuale volta alla realizzazione di un testo compresso.

L'estrazione è un procedimento per cui l'*information designer* seleziona alcuni elementi di un testo che ritiene fondamentali, e li ripropone in sequenza, fornendo così una versione del testo che può appunto definirsi 'compressa'. Non c'è perdita informazionale, perché comunque si fornisce la collocazione di ogni brano estratto, garantendo pertanto al lettore la possibilità di recuperarlo nel contesto originario; o meglio, per usare una terminologia più rigorosa, nel contesto originario.¹⁸⁰ Nel nostro caso, tale soluzione permetteva di risparmiare le differenze espressive dei testi 'estratti', fornendo al tempo stesso al lettore un volume piuttosto breve, da consultare per farsi un'idea sulla metodologia di Tonfoni. Restava da risolvere il problema più interessante per questo lavoro di tesi, quello del reperimento dell'informazione. I curatori del testo, infatti, nel caso in questione, altro non erano che 'motori' umani di ricerca, che leggevano i testi base e ne rintracciavano gli elementi considerati semanticamente rilevanti: svolgevano cioè quell'attività che per il momento è di fatto preclusa all'elaboratore. C'è però una attività in cui il computer resta insuperabile: la memorizzazione dell'informazione. Pertanto si è considerata l'idea di affiancare al libro cartaceo un supporto elettronico, un CD-ROM, sul quale fossero memorizzati completamente i quattro volumi originari. Il computer, insomma, è stato considerato, come suggeriscono Cazalè Bérard e Mordenti, un modello *scriptorium*, che del suo predecessore medievale conserva le

¹⁸⁰ Nel *Trattato* (Umberto Eco, *Trattato di semiotica generale*, Milano, Bompiani, 1975, p.152-161) Eco definisce 'circostanza' la situazione esterna in cui un'espressione, insieme al suo contesto, può occorrere; il contesto è invece l'ambiente nel quale l'espressione occorre insieme ad altre espressioni appartenenti allo stesso sistema di segni. In seguito (Umberto Eco, *Lector in fabula*, Milano, Bompiani, 1979, p.13-19) Eco definirà il contesto una serie possibili di testi ideali di cui si può prevedere l'occorrenza in connessione ad una data espressione, ed il contesto l'ambiente effettivo in cui un'espressione occorre durante un processo comunicativo. Si veda inoltre Umberto Eco, *I limiti dell'interpretazione*, Milano, Bompiani, 1991, p.268.

principali caratteristiche: la conservazione e la ri-scrittura¹⁸¹.

Dopo l'ampio spazio dedicato in questo lavoro all'ipertesto e alle sue funzionalità per *l'information retrieval*, ci si potrebbe domandare come mai non sia stata utilizzata tale strategia comunicativa per memorizzare i contenuti dei quattro volumi. La ragione è semplice: come accennato in precedenza, il computer è sì strumento di conservazione; ma è anche, e soprattutto, strumento di riscrittura. Come ci pare ormai di aver dimostrato, tra le scritture computerizzate possibili, l'ipertesto non è certo la più semplice da realizzare. L'ipertesto richiede infatti tecniche di scrittura particolari, elaborazione ragionata di percorsi di lettura, ridondanza necessaria affinché il lettore non perda mai l'informazione rilevante. Insomma, la forma ipertestuale avrebbe richiesto una riscrittura complessiva dei quattro lavori, pensati, pur nella loro originalità e nella loro ricchezza mediale, per una lettura lineare. Riproporre tutti i testi in modalità ipertestuale non sarebbe stato impossibile, sarebbe anzi realizzabile in futuro, a patto di essere disposti a ripensare completamente la struttura dei quattro volumi. Chi conosce davvero la scrittura ipertestuale, ci sia consentito suggerire, ne conosce anche le difficoltà realizzative, e il rispetto per questa tecnica porta a sconsigliarne l'utilizzo quando non si sia disposti a pagarne i 'costi'. Riversare un testo lineare su CD-ROM, proponendo una pagina piena di *link* al posto dell'indice, e inserendo *hotword* tutte le volte che il testo (concepito originariamente per la carta) rimanda ad un'altra sezione non è un buon servizio che si offre al lettore. Ecco perché si è preferito fornire al lettore un CD-ROM che presentasse i quattro volumi nella stessa identica forma in cui essi si presentano in carta. Il CD-ROM in questione, insomma, si prefigge di essere un comodo sostituto ai quattro volumi originari, molto più leggero, trasportabile, di rapida consultazione: ma non offre informazioni o strumenti di ricerca superiori a quelli che fornirebbero i quattro volumi originari. Quanto detto non è valido ovviamente per il quarto dei volumi, *Abitare il Testo*, che è già stato scritto in maniera ipertestuale, e che pertanto nella veste elettronica ha sfruttato pienamente queste sue caratteristiche.

Il lettore del testo, pertanto, si trova di fronte ad una selezione di brani completamente 'decontestualizzati'; la lettura è più veloce ma più faticosa, inutile negarlo: mancano le ridondanze, gli esempi e i connettivi che facilitano l'apprendimento. Tuttavia in questo modo il lettore può sfruttare

¹⁸¹ **Claude Cazalé Bérard, Raul Mordenti** "La costituzione del testo e la «comunità degli interpreti»", in Patrizia Bellman Nerozzi (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*, Milano, Mimesis, 1997, p.25.

l'operazione di *information retrieval* che altri hanno fatto per lui, recuperare il brano che reputa più interessante per i suoi fini, e andarselo a leggere su CD-ROM nel suo contesto d'origine. Nulla gli impedisce poi di stamparsi questa sezione e leggersela comodamente su carta, oppure di fare un selezione di brani diversa da quella suggerita, alternativa, andandosi a leggere l'opera completa e selezionando i brani ritenuti più rilevanti.

Mi permetto di aprire una piccola parentesi aneddotica collegata alla questione. Spesso tra studenti universitari capita di scambiarsi i libri, le fotocopie, o di rivenderle dopo averli studiati. L'acquirente del testo di seconda mano si ritrova di fronte un testo inevitabilmente 'vissuto', che presenta annotazioni, cancellature, e spesso sottolineature. Ora, la mia esperienza universitaria mi ha insegnato che studiare su un testo colmo di sottolineature è poco proficuo, quando non addirittura fastidioso: sottolineare tutto è come non sottolineare affatto¹⁸². Viceversa, se il lettore che ci ha preceduti è stato accorto nella sua pratica di studio e ha sottolineato solo alcuni periodi per ogni pagina, quella sottolineatura può rivelarsi preziosa. In casi particolarmente disperati, c'è chi si affida completamente al lavoro del lettore precedente, e studia soltanto le sottolineature. Un esempio sconsigliabile, non c'è dubbio. Ma c'è anche chi fa buon uso di quelle sottolineature, le considera una interpretazione magari discutibile, ma frutto comunque di un'operazione selettiva; e magari confronta quell'interpretazione con la sua. Insomma, al contrario del libraio, che valuta il valore di un libro di seconda mano a seconda dell'integrità delle pagine, per uno studente spesso un libro sottolineato con criterio vale il doppio. Ebbene, il lavoro di compressione testuale realizzato si propone come una versione raffinata e meditata dello stesso processo.

In questo paragrafo sono state accennate parecchie questioni rilevanti che meritano un ulteriore approfondimento. Risulta opportuno allora procedere all'analisi di quanto è avvenuto nel caso in questione, seguendo un criterio cronologico. Ci soffermeremo a riflettere sulle questioni qui sollevate ma non approfondite tutte le volte che se ne presenterà l'occasione. Cominceremo con l'occuparci della compressione testuale e del processo¹⁸³

¹⁸² Ci 'gloriamo' di affermare che dello stesso parere è Eco, il quale afferma: "La sottolineatura personalizza il libro. Segna le tracce del vostro interesse. Vi permette di ritornare a quel libro anche dopo molto tempo ritrovando a colpo d'occhio quello che vi aveva interessato. Ma bisogna sottolineare con criterio. Ci sono coloro che sottolineano tutto. È come non sottolineare nulla." **Umberto Eco**, *Come si fa una tesi di laurea*, Milano, Bompiani, 1977, p.137.

¹⁸³ Il termine processo ha numerosi significati che possono talvolta creare qualche fraintendimento. Tonfoni, in *Communication Patterns and Textual Forms*, p.11, definisce il processo "a dynamic set of

che ne è alla base, l'estrazione testuale. Poi analizzeremo il ruolo interpretativo del 'selezionatore' (in inglese tale ruolo si definisce *abridger*), osservando i processi cognitivi che hanno caratterizzato i protagonisti di questa esperienza. Infine prenderemo in considerazione la realizzazione del progetto, analizzandone l'utilizzo di alcuni espedienti grafici e lo sfruttamento (o meno) dei supporti elettronici.

4.2. La compressione testuale

Le procedure atte a realizzare la compressione testuale che sono state seguite nel corso della produzione del libro sono quelle elaborate dagli studi di Tonfoni, relative alla metodologia CPP-TRS¹⁸⁴.

La teoria della compressione testuale può essere considerata una versione micro¹⁸⁵ della macroteoria della comprensione testuale. Laddove quest'ultima, infatti, si concentra sulle potenzialità del testo ancora *in nuce*, cioè sulle possibili diverse realizzazioni espressive di uno stesso testo, la teoria della compressione testuale si applica sul testo già attualizzato e definitivamente completato. In effetti, il nostro lavoro di compressione testuale è partito da quattro testi già completi e definiti. La compressione testuale si definisce 'quantitativa' perché riduce la quantità materiale del testo sforzandosi di mantenerne intatta la qualità. Con il termine compressione quantitativa in CPP-TRS infatti si indica un processo di sintesi molto 'spinto' che si verifica senza perdita di informazione, ma che non esclude l'inevitabile 'cambiamento' qualitativo del testo:

"To compress" means literally "to press together and force into less space"; a compressed text is a text in which information density is *being* reorganized not just quantitatively but qualitatively with a multilayered distribution and in different textual space.¹⁸⁶

L'informazione non si disperde, ma si dispone diversamente. Perché il testo conservi la stessa informazione in minor spazio, deve essere pensata

cognitive operations resulting in linguistic facts and events and communicative actions".

¹⁸⁴ Per questo paragrafo, si è fatto ampiamente riferimento al sito web di Tonfoni relativo alla metodologia CPP-TRS. [Http://www.cirfid.unibo.it/project/ile](http://www.cirfid.unibo.it/project/ile)

¹⁸⁵ Attribuendo a questo termine il significato riconosciuto dalla scienza sociale, e cioè di approccio che approfondisce ed analizza un settore ben definito tenendo presenti i dettami della corrispettiva teoria 'macro', che invece offre un punto di vista più generale e meno specifico.

¹⁸⁶ **Graziella Tonfoni**, *Communication Patterns and Textual Forms*, Exeter, Intellect, 1996, p. 39.

una riorganizzazione complessiva di tale spazio, e non una semplice riduzione. Per comprendere questo fenomeno bisogna tenere presenti alcune peculiari caratteristiche del linguaggio naturale: i segmenti testuali, le unità che lo compongono possono essere compresse a livello locale o globale, tramite un'operazione sintetica, o possono essere espanse in caso di procedimento analitico; inoltre ci può essere astrazione, cioè la condensazione di un testo in un concetto o in una serie di concetti. Tutto ciò non può che cambiare la natura del testo stesso, che pur conservando lo stesso contenuto di base, può generare diverse forme di comprensione ed interpretazione.

Per agevolare il lettore, durante l'operazione di compressione testuale, che in questo caso si realizza attraverso l'estrazione di stringhe letterali di testo, si scelgono frasi sintatticamente semplici, che possono ripetere anche più volte lo stesso concetto ripreso da fonti diverse: in questo modo l'informatività resta comunque alta, perché si offrono sfumature differenti del medesimo pensiero, e al tempo stesso se ne facilita l'apprendimento. Inoltre, come vedremo in seguito, nella nostra esperienza il lettore è stato estremamente facilitato dall'utilizzo di icone che rappresentano gli acronimi dei titoli dei volumi da cui i pacchetti sono stati estratti.

Sia ben chiaro che per quanto 'spinta' possa essere la sinteticità che ci si prefigge, essa non giustifica affatto la perdita della struttura sintattica della frase, come ad esempio avviene se in alcuni casi si elimina il verbo: una forma di questo tipo è quella che si usa a volte per prendere appunti. Nel caso della compressione quantitativa il rispetto della sintassi risulta basilare, perché, quest'ultima, in quanto tessuto fondamentale della comprensione, agevola il lettore nell'interpretazione del testo. I ritocchi da attuare, se davvero indispensabili, devono essere minimi.

La prima fase di un processo di compressione testuale basato sull'estrazione è caratterizzata dalla definizione dei territori testuali su cui agire. In seguito si procede alla selezione dei brani, effettuando un continuo lavoro di controllo. Tale controllo deve essere autonomo: non è cioè opportuno passare all'unità testuale successiva finché non si è raggiunta la completezza nella parte precedente. Le tracce di tali operazioni possono essere conservate, per essere fornite in un secondo momento a chi fosse interessato a vedere come è avvenuta l'operazione di compressione; normalmente, tuttavia, solo il risultato finale di tutto il processo è visibile.

La compressione testuale per realizzarsi validamente richiede che il

materiale testuale da comprimere sia stabile, cioè bene definito e delimitato. Inoltre essa deve essere applicata a dati il più possibile oggettivi. Una selezione testuale, e la relativa estrazione di contenuti, è infatti un'operazione, non ci stancheremo mai di ripeterlo, assolutamente soggettiva e arbitraria, per quanto basata su un lavoro scientificamente giustificabile. Pertanto un'eventuale applicazione a testi di commento, per esempio, potrebbe produrre risultati che distorcono l'informazione, dando così l'impressione di aver applicato una tecnica manipolatoria. Ecco perché i testi ideali su cui operare sono proprio i manuali di istruzioni e la manualistica in genere.

Talvolta si può decidere di intervenire sul testo per collegare i vari estratti e renderne più fruibile la lettura: è il caso per esempio dell'operazione svolta da Richardson¹⁸⁷, che si differenzia dal lavoro da noi analizzato proprio perché predilige la facilità di lettura del testo finale a discapito del rispetto della struttura sintattica del lavoro originale. Si tratta di una scelta che può essere attuata con successo: l'importante, prima di prendere qualsiasi decisione sulla modalità di compressione da realizzare, è che l'*abridger* analizzi con cura il materiale su cui dovrà lavorare e il pubblico a cui rivolgersi, agendo di conseguenza.

4.3. Il ruolo dell'abridger

Tornando al lavoro preso in esame, esso implica l'uso massiccio dell'autoriflessione e dell'autocritica, come unica modalità per garantire l'impacchettamento appropriato dell'informazione: in casi come questo, è cioè necessario vagliare ogni frase dal punto di vista del futuro fruitore. Nel nostro caso, l'autoriflessione è stata favorita sia dal fatto che i tre curatori erano seguiti comunque dall'autrice, sia dalle conoscenze che essi, in maniera diversa, avevano già acquisito sulle tecniche di compressione. Non è stato difficile pertanto schematizzare i processi cognitivi attraverso i quali l'estrazione testuale è stata realizzata, e l'esperienza dei curatori ha limitato i rischi prodotti dalla cosiddetta *protocol analysis*, i rischi cioè di realizzare un modello cognitivo poco rispondente alla realtà. Come già accennato in precedenza, questa tecnica deduce i processi che si realizzano nella mente di chi svolge una attività cognitiva complessa, quale per esempio scrivere, attraverso una serie di domande poste dal ricercatore al soggetto osservato. Per essere più precisi, un protocollo è "a description of the activities, ordered

¹⁸⁷ **Graziella Tonfoni**, *Writing As a Visual Art*, Intellect, Exeter, 1994 (con J.E. Richardson revisore tecnico, e una prefazione di Marvin Minsky; seconda edizione 2000).

in time, which a subject engages in while performing a task”¹⁸⁸. Non basta dire che un protocollo è una descrizione, perché quest’ultima potrebbe interessarsi esclusivamente dei risultati finali di una attività, o dei suoi obiettivi, mentre il protocollo segue passo dopo passo l’attività, comprese le fasi di stallo o i percorsi inutili che solitamente svolgiamo prima di risolvere un problema. Hayes e Flower sottolineano il loro interesse nei confronti della sequenza di azioni che portano alla soluzione di un problema. Tipicamente, una *protocol analysis* si realizza ponendo un problema al soggetto osservato, per esempio un problema aritmetico che richieda un certo ragionamento logico; l’osservatore chiede continuamente al soggetto a cosa sta pensando e cosa ha intenzione di fare, facendo bene attenzione a non influenzare, anche semplicemente con il tono della voce, le scelte di quest’ultimo. Più il numero di soggetti osservati è alto, più è plausibile la realizzazione del modello che lo studioso concepisce alla fine. Tuttavia, spesso i protocolli sono incompleti, perché i soggetti non sempre sono in grado di esprimere verbalmente i loro pensieri, e vanno completati con l’esperienza e l’intuito del ricercatore.

The psychologist’s task in analyzing a protocol is to take the incomplete record that the protocol provides together with his knowledge of the underlying psychological processes by which the subject performs the task(...) Between surfacings, the mental process, like the porpoise, runs deep and silent. Our task is to infer the course of the process from these brief traces.¹⁸⁹

Tornando al nostro lavoro, è stato realizzato un modello che non può definirsi il frutto di una *protocol analysis* ortodossa, ma che presenta alcuni aspetti in comune con questa tecnica. Innanzitutto, la presenza di un processo di *problem-solving*; in questo caso il problema da risolvere era l’estrazione di porzioni di testo rilevanti; poi, la presenza di un esperto, cioè di una figura-guida dell’analisi, volta in un certo qual modo a monitorare il processo analitico. Le due principali differenze, invece, erano le seguenti: innanzitutto, i ‘soggetti’ analizzati, i tre curatori, non erano ingenuamente all’oscuro dei fini dell’analizzatore, come di solito accade, ma erano almeno in parte consci dei processi che realizzavano; in secondo luogo, l’osservazione non si

¹⁸⁸ John R. Hayes, Linda S. Flower, “The Dynamics of Composing”, in Gregg Lee W., Steinberg E.R. (eds.), *Cognitive Processes in Writing*, Hillsdale, New Jersey, Erlbaum 1980, p.4.

¹⁸⁹ John R. Hayes, Linda S. Flower, op.cit.,p.9.

verificava durante il processo stesso, ma a posteriori. I tre curatori del testo, infatti, guidati dall'autrice stessa, in questo caso anche con il ruolo di ricercatrice, al termine del lavoro si sono interrogati sulle fasi che si sono succedute durante l'estrazione dei brani. Come vedremo, esse non sono identiche: i profili cognitivi che ne emergono sono differenti.

Da una parte, ciò può essere dovuto all'esiguo numero di casi presi in considerazione, tre appunto, e alle diverse porzioni di testo che i tre curatori comprimevano; dall'altra, potrebbe indicare alcune differenze nella realizzazione di processi cognitivi dovute a diversi retroterra culturali, a diversità personali acquisite quali metodo di studio e tecnica di ricerca, o innate; in ogni caso tendenze cognitive precise. Vediamo lo schema che modella il processo di estrazione testuale realizzato:

Modello di estrazione n°1 Carmine Caputo	Modello di estrazione n°2 Michele Semenza	Modello di estrazione n°3 Eduardo Ayala
1. Riferimento ad abilità accertate e consuetudini consolidate 2. Direzione concettuale e pre-impressionamento testuale 3. Lettura a percezione massima e assorbimento 4. Individuazione frase massimamente tematica 5. Allargamento area tematica 6. Prioritarizzazione sottoaree tematiche 7. Identificazione e selezione sottoarea/e massimamente rilevanti	1. Riferimento ad abilità accertate e consuetudini consolidate 2. Ricaricamento conoscenze precedenti 3. Lettura reattiva a percezione media del materiale 4. Individuazione aree tematiche massimamente rilevanti 5. Selezione sequenze massimamente rappresentative 6. Valutazione qualitativa e quantitativa	1. Riferimento ad abilità accertate e consuetudini consolidate 2. Ricaricamento lingua 2 3. Ricaricamento conoscenze precedenti 4. Lettura reattiva a percezione media del materiale 5. Individuazione aree tematiche massimamente rilevanti 6. Selezione sequenze massimamente rappresentative 7. Valutazione qualitativa e quantitativa
GT/AT → AT → SM → CAV → [PSM]	GT/AT → AT → SM ↘ PSM	GT/AT → AT → SM ↘ PSM

Tabella 3: tre modelli cognitivi di estrazione testuale

Prima di procedere con alcune riflessioni su questo modello, occorre rivolgere l'attenzione su due aspetti sinora trascurati, ma di fondamentale rilevanza: le differenze 'personali' tra i tre curatori, che per comodità d'ora in avanti chiameremo MOD1, MOD2 e MOD3, e i criteri di divisione del materiale tra gli stessi.

MOD1 è il modello che riguarda il sottoscritto. Nel periodo in cui il progetto è stato realizzato, ero uno studente che frequentava il corso di Linguistica Computazionale della professoressa Tonfoni, con assai scarse conoscenze della metodologia esposta nei quattro testi 'originari', anche perché il corso di quell'anno non aveva come oggetto la metodologia e, conseguentemente, i volumi in questione. Questa mia 'ignoranza' non è stata considerata, tuttavia, una condizione svantaggiosa, anzi; come si diceva in precedenza, l'*information designer* non è tenuto ad essere un esperto del settore trattato. Talvolta questa eventuale 'esperienza' può essere negativa e influenzare le scelte in maniera sbagliata, specie nel caso della estrazione testuale. Il *designer*, infatti, deve porsi in una condizione quanto più possibile vicina a quella del lettore, il quale non è affatto un esperto, ma vuole saperne di più. Chiaramente, pur non conoscendo la metodologia, conoscevo le tecniche di estrazione testuale che la professoressa aveva ampiamente illustrato durante le lezioni. Deve essere ricordato, di nuovo, che l'esperienza è nata inizialmente come esercitazione didattica, e solo in seguito ha trovato una realizzazione editoriale al di fuori dell'ambito accademico.

Per quanto concerne MOD2, la condizione di partenza è quasi diametralmente opposta alla mia. Il curatore in questione, al momento dell'estrazione testuale, era infatti in una situazione di più che discreta conoscenza della metodologia, che aveva già studiato approfonditamente e sulla cui applicazione nella didattica supportata a distanza stava preparando la tesi di laurea.

Ancora diversa, infine, la condizione di MOD3, per molti aspetti affine a quella di MOD2. Anche in questo caso, infatti, il curatore conosceva bene la metodologia esposta nei quattro volumi. In particolare, conosceva molto bene la sezione relativa alle tele testuali¹⁹⁰, sulle quali stava preparando la tesi di

¹⁹⁰ Per chi volesse approfondire la conoscenza delle tele testuali, talvolta definite 'schemi', tra i testi di Graziella Tonfoni, quelli che trattano più approfonditamente l'argomento sono senz'altro *Scrittura multimediale*, Treviso, Pagus, 1991, e *La comunicazione aziendale come arte visiva*, Treviso, Pagus, 1991.

dottorato. Il suo era il caso di una conoscenza talmente ‘profonda’ da poter divenire controproducente ai fini dell’operazione; pertanto, la soluzione stabilita d’accordo con gli altri, come vedremo in seguito, fu quella di non far trattare affatto, a lui, le sezioni dei volumi dedicate alle tele testuali. Una caratteristica propria di MOD3, che lo differenziava tanto da MOD1 quanto da MOD2, era la lingua madre; nel caso in questione, infatti, l’operazione di estrazione era influenzata dal fatto che il curatore fosse di madrelingua spagnola, benché conoscesse perfettamente l’italiano, la lingua dei volumi da comprimere.

Queste dunque le differenze tra i tre curatori, che come vedremo, potrebbero aver avuto non poca influenza sui rispettivi processi di estrazione testuale. Dati questi punti di partenza, si stabilì una divisione del lavoro. Le ipotesi erano varie: si poteva affidare a ciascuno l’incarico di comprimere un libro. Oppure si poteva fare comprimere da tutti e tre tutti i volumi, operando poi una scelta, volta per volta, tra i risultati delle selezioni considerati migliori, analizzando per esempio i passi selezionati come rilevanti da tutti rispetto a quelli selezionati da un solo curatore. Oppure si poteva lasciare che a decidere quale delle selezioni preferire fosse l’autrice del testo originario. La prima ipotesi fu scartata immediatamente, perché legata più alla fisicità dell’oggetto libro che ai suoi contenuti. Se si considera quanto detto a riguardo dei quattro testi, in relazione alle tipologie di pubblico diverse a cui erano indirizzati, delle diversità di stile, delle ridondanze ‘intertestuali’ presenti, si comprende come non fosse una buona idea comprimere separatamente i testi, perpetuando così la separazione fra di loro. Il volume finale doveva essere complessivamente coerente e ci si doveva sforzare di amalgamare il più possibile i contenuti, non poteva presentarsi come un semplice indice ‘allargato’ non solo ai titoli delle sezioni ma anche ad una parte di esse. Per cui, se una divisione del lavoro doveva esserci, il criterio non poteva essere il semplice ‘un libro a testa’. La seconda ipotesi fu scartata perché l’estrazione testuale non può risolversi come mera intersezione statistica di più operazioni. Nel momento in cui un curatore seleziona le frasi A e B, questa scelta probabilmente influenzerà la successiva selezione della frase C; per questa ragione, o si prende un testo compresso nella sua interezza dalla stessa persona, che ha usato gli stessi principi euristici e le corrispondenti tecniche cognitive, oppure si rischia di fare un *collage* molto poco efficace. L’ultima possibilità, quella di lasciare la scelta finale all’autrice, fu scartata sia perché stabiliva una competitività all’interno del gruppo di lavoro che avrebbe potuto comportare tensioni, che alla fine

avrebbero nuociuto al risultato finale, sia perché la presenza dell'autore che segue il lavoro dell'*information designer* è più un'eccezione che la norma. In effetti la natura didattica del nostro lavoro faceva sì che l'autore (in questo caso autrice) dei testi originari fosse disponibile per confronti o scambi di idee; se viceversa l'esperienza fosse nata e si fosse sviluppata unicamente in ambiente editoriale, probabilmente l'autore si sarebbe fatto presente solo di fronte al lavoro terminato, e non prima. E poi, porre l'autore di fronte ad una scelta tra alternative, equivale per certi versi a chiedergli di fare egli stesso il lavoro di compressione; allora si sarebbe trattato comunque di un'altra operazione.

La scelta adottata fu pertanto quella di dividerci il lavoro, operando però una divisione sui contenuti della metodologia indicati 'a monte' come rilevanti dall'autrice: MOD1 si sarebbe occupato delle tele testuali, MOD2 degli oggetti testuali e di parte delle macchine testuali, MOD3 di segnali e simboli oltre che delle rimanenti macchine. Successivamente, quando si decise di dare un seguito all'esercitazione realizzando un vero e proprio prodotto editoriale, MOD1 e MOD3 si sono occupati di prefazione e introduzione, mentre MOD2 ha realizzato la veste grafica e la bibliografia; ma questo è un altro discorso che per il momento non ci interessa. La professoressa Tonfoni, nel duplice ruolo di docente e di autrice, sarebbe intervenuta nella prima fase, quella di preparazione all'esercitazione, in qualità di docente; avrebbe poi lasciato ampia libertà ai tre curatori, in contatto fra loro, per la realizzazione del lavoro; sarebbe nuovamente intervenuta alla fine, stavolta nelle vesti di autrice, per accertare la qualità del testo compresso e la coerenza rispetto ai testi originari. Nei fatti l'attività si è svolta proprio così come era stata progettata.

Riprendiamo adesso lo schema raffigurato in tabella 3, e in particolare la parte inferiore, che rappresenta le priorità seguite dai tre curatori nei loro 'movimenti di estrazione' tra le fonti testuali e non. In essa appaiono le sigle dei quattro volumi originari (SM sta per *Scrittura multimediale*, CAV per *La comunicazione aziendale come arte visiva*, PSM per *Partitura solfeggio movimento*, AT per *Abitare il testo*), più una quinta sigla (GT) che rappresenta il ruolo 'propulsivo' ricoperto dall'autrice. Poiché la preparazione è stata uguale per tutti, l'inizio dell'attività di ricerca, per i tre curatori, è stata comune, e si è estesa attraverso l'utilizzo degli appunti delle lezioni e mediante i consigli della docente che suggeriva le aree sulle quali concentrarsi di più e quelle maggiormente trascurabili, senza per altro porre

vincoli di qualsiasi natura. Immediatamente dopo i suggerimenti della professoressa, sia cronologicamente che gerarchicamente, ha trovato spazio la lettura e la compressione di *Abitare il Testo*. La ragione è piuttosto semplice: si tratta di un testo che, per buona parte, è già una sintesi piuttosto spinta dei tre precedenti volumi. Pertanto leggere immediatamente *Abitare il testo* poteva offrire ai curatori un quadro piuttosto esauriente della struttura complessiva che spettava loro ricostruire con i tasselli ‘prelevati’ dai testi originari. Ma *Abitare il Testo*, per la sua intrinseca natura, non solo si è rivelato un buon punto di partenza per dirigere lo sguardo sul lavoro da farsi, ma è stato anche il primo testo ad essere compresso. In effetti la prima frase estratta che compare nella versione definitiva dell’opera è quasi sempre proveniente da *Abitare il Testo*, che definisce gli aspetti salienti della metodologia con chiarezza e immediatezza, anche se talvolta in maniera molto succinta. In aiuto sono venute allora le frasi estratte dagli altri testi che approfondiscono e ampliano il discorso introdotto da *Abitare il Testo*.

Dopo questa prima fase comune, le strade dei tre curatori si sono ovviamente separate, come era stato previsto già nella fase di pianificazione. MOD1, si è detto, si è occupato delle tele testuali, e pertanto si è concentrato sui testi che trattavano questo argomento: prima di tutto *Scrittura multimediale*, per passare poi in maniera lineare a *La comunicazione aziendale come arte visiva*. In secondo piano è rimasto il terzo volume della serie, *Partitura solfeggio movimento*, da cui non sono state estratte informazioni, dal momento che il testo non tratta l’argomento delle tele, ma è stato comunque letto per ‘completezza’. Si ricordi a questo proposito che MOD1 non conosceva la metodologia illustrata nei testi, diversamente dagli altri due curatori; e che pertanto sentiva maggiormente l’esigenza di una lettura completa, affinché nulla di rilevante fosse scartato. Per quanto riguarda MOD2 e MOD3, la parte inferiore dello schema riportato in tabella 3 evidenzia bene come i due curatori abbiano seguito un percorso di ricerca tra le fonti identico: partenza sempre basata su *Abitare il testo*, quindi sviluppo dell’estrazione ‘in parallelo’ tra *Scrittura multimediale* e *Partitura solfeggio movimento*, mentre è stato trascurato *La comunicazione aziendale come arte visiva*, e questo è abbastanza ovvio, dal momento che questo testo non interessava i contenuti metodologici compressi dai due curatori.

È possibile fare qualche riflessione già sul diverso cammino di ricerca tra le fonti realizzato tra MOD1 e gli altri due modelli. In particolare, laddove

MOD2 e MOD3 operavano una selezione tra due testi ponendoli sullo stesso piano (in un circuito elettrico si parlerebbe appunto di collegamento in parallelo: i due elementi sono alternativi e non esiste gerarchia fra di essi) MOD1 realizzava una selezione ‘in serie’, cioè sceglieva prima i brani dal testo 1, e poi quelli dal testo 2, senza cambiare quasi mai quest’ordine. I due testi si presentavano dunque come non sostituibili, ma gerarchicamente organizzati (il testo 1 è indispensabile al testo 2, proprio come in un circuito in serie). Si è detto senza cambiare *quasi* mai quest’ordine, perché in determinati casi quest’ordine veniva comunque modificato: si rammenti che questo schema è frutto di una generalizzazione, che cerca induttivamente di stabilire una regola partendo da alcuni casi. Ma tale regola, in quanto tale, deve essere in grado di accettare anche le inevitabili eccezioni, non va considerata come rigida e applicabile in qualsiasi contesto, altrimenti si rischia di forzare il modello inaridendone le capacità esplicative. Detto ciò, possiamo concludere che con ogni probabilità i diversi percorsi di ricerca testuale tra le fonti realizzati tra i curatori sono attribuibili soprattutto a due fattori: la diversa natura dei testi, e la diversa preparazione iniziale, di cui si è già detto, per cui MOD1 si è sentito tranquillizzato dal confronto con tutti i testi anche se era possibile escluderne alcuni a priori in base ai suggerimenti dell’autrice, cosa che MOD2 e MOD3 hanno fatto.

MOD1 ha trovato comodo ed efficace, nella sua stesura del testo compresso, partire sempre da *Scrittura Multimediale*, perché questo testo è il meno sintetico e il più ridondante, dati i suoi obiettivi didattici e il pubblico di formatori a cui è indirizzato, passando poi a *La comunicazione aziendale come arte visiva*, che è senz’altro più sintetico ma anche meno completo e più specificamente indirizzato ad un pubblico selezionato, ‘di nicchia’ direbbero gli economisti, quale appunto quello degli operatori aziendali. La scelta di MOD1 è stata quella di operare dapprima sul corpus più sintetico in assoluto, quello costituito da *Abitare il testo*, per avere un quadro complessivo; di procedere quindi sul corpus più ampio e completo, quello di *Scrittura multimediale*, per essere sicuro di non perdere nessun aspetto rilevante; di passare infine a *La comunicazione aziendale come arte visiva* per rimarcare gli aspetti più interessanti già estratti da *Scrittura multimediale*. Si tratta di un procedimento di reperimento dell’informazione che nel mio caso si è rivelato molto efficace, ma che non è generalizzabile a meno che non si verificano le stesse condizioni che si sono realizzate per il mio lavoro. E che comunque ci ricorda, se fosse ancora necessario, che i processi di ricerca devono sempre e comunque adattarsi all’ambiente testuale

in cui si realizzano, per sfruttare a pieno le loro potenzialità. Infatti lo stesso procedimento non è stato adottato da MOD2 e MOD3, i quali si ponevano di fronte a due testi che presentavano una certa omogeneità di stile, e i cui legami erano semmai cronologici; ragion per cui, era preferibile scegliere di volta in volta quali frasi estrarre e da quale testo, anche perché il secondo, *Partitura solfeggio movimento*, essendo posteriore di un anno rispetto al precedente, poteva presentare alcune evoluzioni della metodologia ancora inesprese in *Scrittura multimediale*.

Veniamo ora alla prima parte dello schema presentato in tabella 3, quella superiore. Graficamente si sono posti sullo stesso piano orizzontale i processi che presentavano aspetti simili, o per lo meno somiglianti in parte. Quando un processo è stato attuato esclusivamente da uno dei tre curatori, lo spazio sulla riga corrispondente agli altri due è stato volutamente lasciato vuoto. In questo modo si è cercato di evidenziare i diversi percorsi che hanno portato i curatori allo stesso risultato finale, cioè un testo compresso. Le fasi di ogni processo sono state poi ordinate in maniera sequenziale, così che, per fare un esempio, la stessa attività, o comunque un'attività con numerosi aspetti comuni in tutti e tre i processi, per MOD1 può essere la settima tappa del percorso, in MOD2 la quinta, in MOD3 la sesta. Non bisogna perciò commettere l'errore di confrontare le varie fasi basandosi sulla cifra che le precede, perché questa indica la loro posizione interna nel processo, e non una posizione assoluta, dal momento che ogni processo può essere costituito da un numero diverso di fasi.

Il cammino di estrazione testuale dei tre curatori comincia con la stessa procedura, e cioè il riferimento ad abilità accertate e consuetudini consolidate. Ogni persona che si ponga alla ricerca dell'informazione, infatti, lo fa presupponendo una serie di tecniche abitudinarie delle quali non potrebbe sbarazzarsi neanche se lo volesse, né di fatto deve volerlo. Innanzitutto, è ovvio, bisogna saper leggere; ma ci sono diversi modi di leggere. C'è chi legge velocemente e chi lo fa con maggiore lentezza; chi preferisce leggere ad alta voce per memorizzare più a fondo i concetti; chi prima di leggere sfoglia il testo nella sua completezza. Tutte queste abitudini fanno parte di un bagaglio personale che riflette inevitabilmente la formazione ricevuta, ma anche i gusti di ciascuno. Pertanto, da un punto di vista didattico, l'unico consiglio che il formatore potrebbe dare all'apprendista 'ricercatore d'informazione' è, secondo la nostra opinione, quello di sforzarsi innanzitutto di rendersi consapevoli di queste abitudini.

Solo in questo modo, infatti, sarà possibile sfruttarle al meglio quando è il caso, e cercare di correggerle quando sono giudicate poco efficaci: ancora una volta, la parola d'ordine è metacognizione, cioè pensare al come cognitivamente operiamo per esserne consapevoli ¹⁹¹

Tornando al nostro schema, la prima fase fa riferimento ad abilità e consuetudini che precedono il testo. Si faccia attenzione alla scelta degli aggettivi utilizzati: abilità accertate e consolidate. La base di partenza deve essere ben solida perché un processo cognitivo sia efficace, perché altrimenti il dominio dell'incertezza renderà insicura qualsiasi scelta. Dopo questo primo passaggio comune, le strade dei tre ricercatori si separano. MOD1, in particolare, realizza un'operazione definita di 'Direzionamento concettuale e pre-impressionamento testuale' che negli altri due è assente. L'operazione consiste in una consultazione superficiale dei volumi, che ne offre una visione generica e poco accurata, in grado però di 'direzionare' i processi successivi attraverso appunto un pre-impressionamento. Si pensi ad un

¹⁹¹ Mi sia concessa un'altra breve divagazione. Qualche tempo fa mi accorsi che senza una ragione precisa trovavo una certa difficoltà nello studio in biblioteca. Pensai che fosse dovuto alle distrazioni presenti; ma il risultato era identico anche in biblioteche deserte. Pensai alle condizioni luminose, alla comodità del banco, a tutta una serie di fattori che potessero influenzare il mio studio, senza ottenere alcun risultato. Improvvisamente, era come se non riuscissi più a concentrarmi in biblioteca: eppure non era sempre stato così. Solo riflettendo a lungo sulle mie modalità di studio trovai la soluzione a questo 'incretinoso' problema. Per circa tre anni e mezzo, infatti, avevo vissuto in un collegio, in una stanza piuttosto piccola e con una scrivania scomodissima, progettata e posizionata in maniera tale da contraddire tutti i concetti base (e non solo quelli) della teoria ergonomica. Ebbene, questa scrivania era talmente scomoda che avevo cominciato a leggere i libri da preparare per l'esame sdraiato sul letto. Dapprima avevo incontrato qualche difficoltà, poi con il passare del tempo la postura era diventata per me talmente comoda che riuscivo a studiare proficuamente soltanto sdraiato. E per quanti servizi possa offrire, una biblioteca non offre ai suoi utenti la possibilità di sdraiarsi. Anche il processo di estrazione testuale, tornando alla strada maestra del nostro discorso, l'ho svolto sdraiato languidamente sul mio letto, benché ormai mi sia trasferito e possa godere di una comodissima scrivania. Ho aperto questa parentesi 'privata' perché mi è servita a riflettere su come certe abitudini di solito siano tenute al di fuori degli studi cognitivi o testuali, perché ritenute non rilevanti, o accantonate in isolati studi 'pragmatici'. Invece tenerne conto è indispensabile per realizzare come alcuni processi si verifichino, e come sia eventualmente possibile modificarli. Per fare un altro esempio, quando si progettano i primi 'libri elettronici', ci si mosse esclusivamente all'interno dell'area squisitamente testuale, e da questo punto di vista, i vantaggi del testo visualizzato da un monitor erano molteplici rispetto a quello stampato su carta. Solo quando ci si rese conto che la gente è abituata ad un contatto 'fisico' con il testo, a portarselo sul divano o in metropolitana, a identificarlo come un 'oggetto chiuso', ci si rese conto che il libro elettronico avrebbe potuto essere complementare, ma non sostituire quello cartaceo. A tal proposito, la ricerca tecnologica non si è lasciata scoraggiare, ed ha previsto una serie di dispositivi di lettura all'apparenza molto simili al libro tradizionale, e come tali facilmente maneggiabili, che altro non sono che piccoli monitor dotati di memoria ed in grado di visualizzare numerose pagine ed immagini. Nel momento in cui si scrive, i progetti più interessanti sono il Rocket eBook™ (<http://www.rocket-ebook.com>), 650 grammi di peso per uno schermo di di 9x14 cm, e il Softbook™, 1,3 kg per uno schermo di 6x8 cm. Cfr. **Antonio Spadaro**, "Il libro e internet: le tecnologie digitali e il mondo dell'editoria", in *La Civiltà Cattolica*, n°3591, gennaio 2000, p. 33.

contadino che deve arare un campo, e, in sella al suo trattore, scruta l'orizzonte per rendersi conto di quanto ampio è il campo, di quanto tempo più o meno necessiterà, di quale ritmo dovrà seguire per terminare il lavoro in giornata. È ovvio che se l'agricoltore è il proprietario del campo, o ci ha già lavorato, non farà tutto ciò: conoscerà come le sue tasche la terra da arare, e sarà pertanto in grado di orientarsi sul da farsi senza bisogno di questa fase di osservazione. Nel nostro processo di estrazione si è verificata una situazione simile: MOD1 ha ritenuto necessario osservare il suo campo¹⁹² d'azione, prima di cominciare il lavoro, laddove MOD2 e MOD3 hanno ritenuto superficiale fare altrettanto. Si ricordi a questo punto che la realizzazione dello schema è stata posteriore all'attività vera e propria, che è stata svolta indipendentemente dai tre curatori, i quali pertanto non hanno potuto influenzarsi a vicenda sulle scelte operative da compiere.

La fase 2 di MOD3 è esclusiva ed è fortemente contestualizzata: come detto in precedenza, l'autore è di madre lingua spagnola, per cui ha svolto un lavoro aggiuntivo, che è quello costituito dal dover ricorrere all'analisi dei testi in lingua straniera, con il relativo 'bagaglio' di regole sintattiche e patrimonio lessicale, insomma con tutte le conoscenze che una lingua comporta e che non può ripristinare in maniera automatica perchè non gli sono proprie, o, come direbbe una certa sociologia linguistica, 'incarnate'¹⁹³. Il richiamare questo insieme di conoscenze, che quando parliamo nella nostra madre lingua è talmente automatico da essere impercettibile, è stato appunto definito 'ricaricamento in lingua 2', dando per sottinteso il caricamento in lingua 1, cioè in lingua madre, fatto dagli altri due curatori. Va rimarcato come nell'essere umano non sia possibile, al contrario che per l'elaboratore, operare un 'caricamento' completo di nozioni una volta per tutte e in maniera definitiva; di conseguenza, anche se la fase di 'traduzione' dallo spagnolo all'italiano e viceversa è stata rilevata in questa posizione, si può tranquillamente affermare che essa sia cominciata da questo punto, ma non si

¹⁹² Mai un termine fu più appropriato, vista la metafora precedente.

¹⁹³ Sudnow trova un'interessante analogia tra il parlare e il suonare musica improvvisando: così come il musicista, dopo aver appreso il linguaggio musicale attraverso insegnamenti orali o istruzioni 'scritte' (le note), li abbandona per diventare un vero esecutore, così il parlante utilizza i concetti linguistici quali per esempio quelli di sostantivo, preposizione o avverbio solo per analizzare il linguaggio, oppure per apprendere una lingua straniera. Una volta terminato questo apprendimento, è il corpo che torna ad essere il centro del linguaggio parlato, con il significato che emerge ritmicamente mentre portiamo avanti il discorso, 'improvvisando' come i musicisti di jazz. In questo senso il parlare è incarnato, piuttosto che primariamente mentale. Cfr. **David Sudnow**, *Ways of the Hand: The Organization of Improvised Conduct*, Cambridge (Mass.), Harvard University Press, 1978, **David Sudnow**, *Talk's Body: A Mediation between Two Keyboards*, New York, Knopf, 1979.

sia esaurita certo con il punto successivo. Piuttosto, essa si ripercuote e si ripete ricorsivamente durante tutte le fasi del processo, tanto che non è realistico 'isolarla' una volta per tutte in un ambito preciso. D'altronde, si è già detto che l'utilizzo di modelli computazionali è efficace quando permette di evidenziare processi altrimenti invisibili, ma non deve dall'altro lato costringerci a forzare l'elasticità e la fluidità del pensiero umano 'continuo' nella rigidità e nella operazionabilità dell'operato 'discreto' di un computer.

La fase successiva accomuna MOD2, per il quale è la seconda nell'ordine, e MOD3, mentre è assente in MOD1: si tratta dell'operazione di ricaricamento delle conoscenze precedenti. Si ripensi a quando, nella prima parte di questo lavoro, si sottolineava come un processo di ricerca dell'informazione si muova sempre a partire da una certa cognizione del mondo, da un certo bagaglio di conoscenze che può essere insufficiente rispetto ai fini della ricerca, ma tuttavia c'è e deve essere considerato. Nessun utente si pone come completamente ignaro, di fronte all'informazione, al contrario è in grado di formulare immediatamente dei pre-giudizi e delle schematiche assunzioni basandosi sul proprio bagaglio culturale. È quello che è accaduto ai tre curatori, che prima di intraprendere l'estrazione vera e propria hanno inizialmente fatto riferimento alle loro abilità più generali, poi ad un bagaglio di conoscenze strettamente connesso con i contenuti della ricerca, e cioè alle nozioni in loro possesso sulla metodologia CPP-TRS. Si comprende allora perché MOD1, che non disponeva di questo genere di nozioni, non abbia realizzato questa operazione, operazione che hanno invece realizzato gli altri due curatori. A tal proposito, si noti come da questo punto in poi i comportamenti di MOD2 e MOD3 siano perfettamente identici: pertanto, nel corso della nostra disamina, ci riferiremo, in termini di paragone con MOD1, soltanto ad uno di essi, dando per scontata la congruenza del comportamento dell'altro.

Con la terza operazione si comincia il processo di lettura vero e proprio: MOD1 dichiara di effettuare una lettura che è già approfondita, che già scava nei meandri del testo, potendo contare oltre tutto sul pre-impressionamento che ha ricevuto durante la fase 2. MOD2, invece, preferisce procedere con una lettura che è ancora piuttosto superficiale, che nello schema si è definita 'a percezione media'. Fermiamoci un attimo a riflettere sulla lettura come operazione in sé. La lettura, in quanto processo cognitivo di natura complessa, implica l'organizzazione e il coordinamento di altri sottoprocedimenti cognitivi: l'attenzione, la focalizzazione, la selezione e la

memorizzazione.¹⁹⁴ L'attenzione consiste nel dirigere l'attenzione sulla fisicità del testo escludendo gli altri elementi estranei eventualmente presenti nell'ambiente; la focalizzazione invece è data dal concentrarsi su elementi portanti della struttura testuale, avendo preso in considerazione il testo nella sua globalità; la selezione consiste nel valutare le capacità informative degli elementi testuali, trattenendone solo alcuni; la memorizzazione, infine, è la capacità di organizzare le informazioni selezionate posizionandole in un nuovo contesto cognitivo e attribuendo loro delle relazioni di varia natura. Si può di conseguenza sostenere che in MOD1 la focalizzazione è più accentuata e la selezione più spinta che in MOD2, il quale provvederà in seguito a concretizzare determinate intenzioni qui appena accennate. La lettura realizzata da MOD1 può definirsi una lettura ricettiva, o *receptive reading*; lettura lineare e sequenziale, attiva processi di memorizzazione e procede "sintonizzandosi" con l'autore, cioè fa attivare le facoltà di inferenza e di deduzione al fine di ricostruire e interpretare il testo seguendo a ritroso le complesse fasi di produzione elaborate dall'autore¹⁹⁵. Un lettore che si sintonizzi con l'autore si lascia da quest'ultimo guidare nel processo di lettura, seguendo gli elementi testuali selezionati dall'autore per stimolarlo e direzionarlo, ferma restando l'incapacità di qualsiasi autore di controllare completamente la direzione di lettura. Viceversa, la lettura realizzata da MOD2 in questa fase può dirsi una lettura globalmente informativa, o *skimming*. Il lettore ispeziona il testo sul piano globale e localizzando paragrafi e frasi particolarmente interessanti, rimandando a dopo il compito di rivederlo e rileggerlo per eventuali controlli e chiarificazioni.

Il cuore del processo di estrazione comincia con la quarta fase: MOD1 va alla ricerca di una frase particolarmente rilevante, dai forti contenuti tematici, che possa con poche parole sintetizzare un'area semantica ben più vasta. Si tratta, come si noterà, dello stesso procedimento di chi, studiando, sottolinea una frase che lo ha colpito particolarmente, e che gli sembra in grado di 'erigersi sul testo' come punto di riferimento di un settore più vasto. MOD2, invece, individua dapprima un'area rilevante, che può identificarsi, in termini visivi, in un paragrafo, e solo in un secondo momento seleziona le sequenze, all'interno di quest'area, che ritiene maggiormente rilevanti. Mentre il secondo modo di procedere riduce progressivamente l'area testuale, il primo opera scegliendo una frase-chiave e allargando poi l'area d'analisi intorno ad

¹⁹⁴ Graziella Tonfoni, *Sistemi cognitivi complessi*, Treviso, Pagus, 1991, p.67.

¹⁹⁵ Graziella Tonfoni, *op.cit.*, p. 68.

essa. Questo secondo modo di procedere richiede un aggiustamento di tiro: dopo aver allargato l'interesse ad un'intera area tematica, infatti, si rende necessaria una strutturazione gerarchica dell'area stessa, che così com'è è ancora troppo ampia. La fase successiva, definita nello schema "prioritarizzazione sottoaree tematiche", consiste infatti nella disposizione in sottoaree dei passi dell'area di testo selezionata che non sono posti allo stesso livello della frase che ha avuto il ruolo di *trigger*, che cioè ha messo in moto il processo di estrazione, ma che si rivelano tuttavia ugualmente interessanti. Tra questi passi, si procede alla definitiva identificazione e selezione di quelli ritenuti maggiormente rilevanti.

Più apparentemente semplice è il modo di procedere di MOD2, il quale, dopo aver individuato le aree tematiche massimamente rilevanti, procede alla selezione delle sequenze che ritiene più significative. È interessante notare che questo modello di procedimento di estrazione prevede una valutazione finale qualitativa e quantitativa dell'operato, operazione invece assente in MOD1.

Non è nostra intenzione, a questo punto della trattazione, stabilire quale sia il procedimento di estrazione più efficace, anche perché un giudizio valutativo in questo caso dovrebbe partire dai risultati ottenuti e non dalle modalità operate per accedervi. Il nostro interesse è semmai focalizzare l'attenzione sulle diversità esistenti fra questi cammini di estrazione, per mostrare quanto numerosi siano i fattori che influenzano la ricerca dell'informazione: la diversa cultura di base, il diverso bagaglio di conoscenze acquisite sull'argomento, le diverse abitudini personali. Anche le modalità del ragionamento attuato possono dimostrarsi determinanti: pare di poter dire, infatti, che laddove MOD1 proceda attraverso un processo logico vagamente di tipo induttivo, MOD2 (e di conseguenza MOD3) operino un processo di estrazione di marca deduttiva. "Nel caso delle deduzioni logiche c'è una regola, da cui, dato un caso, si inferisce un risultato"¹⁹⁶. Si tratta di un procedimento infallibile, a patto che siano valide le premesse: è il modo in cui ragiona MOD2, il quale stabilisce, nella fase 'Individuazione aree tematiche massimamente rilevanti', che un'area sia di elevato interesse (regola), laddove un'area, non è superfluo ricordarlo, altro non è che un insieme di frasi. Poi individua una frase in quest'area (caso) e ne deduce che anch'essa è interessante, perché fa parte dell'area selezionata (risultato). Si tratta di una modellizzazione imperfetta, ovviamente, perché se MOD2

¹⁹⁶ Umberto Eco, *Trattato di semiotica generale*, Milano, Bompiani, 1975, p.185.

ragionasse in maniera rigidamente deduttiva, allora selezionerebbe tutte le frasi di un'area rilevante, cosa che non è, perché c'è comunque un processo di scelta all'interno dell'area. Tuttavia mi sembra che l'applicazione di un modello deduttivo per analizzare il procedimento di MOD2 regga piuttosto bene, e ci consenta di trarre alcune conclusioni. Questo modo di ragionare fonda la sua validità nelle premesse: se la premessa maggiore (la regola) e la premessa minore (il caso) sono vere, sarà indiscutibilmente vero il risultato. Nel nostro lavoro di ricerca, la fase più delicata è allora la fase 4, che individua le aree: se si individua un'area che non è poi così importante per il valore informativo apportato, anche le frasi selezionate in seguito per deduzione saranno poco informative.

E veniamo adesso al procedimento di MOD1, che abbiamo definito 'vagamente', perché di una similitudine e non di una completa adiacenza si tratta, di tipo induttivo. "Nel caso dell'induzione, dato un caso e un risultato, se ne inferisce la regola"¹⁹⁷. MOD1 non stabilisce a priori quali siano le aree tematiche massimamente rilevanti: piuttosto, va alla ricerca di un caso: una frase che lo colpisca particolarmente per i suoi contenuti. Poi allarga il campo all'intera area, che analizza alla ricerca di altre informazioni interessanti, di riscontri alla sua ipotesi, e cerca di organizzarla seguendo delle priorità. E solo allora induce una regola, e cioè stabilisce che se quest'area contiene una frase molto interessante, allora l'intera area deve essere molto interessante, e dunque conviene effettuare in quest'area il processo di estrazione. L'induzione non è un procedimento logico infallibile, perché induce delle regole che non potranno mai essere dimostrate come certe, ma semmai come altamente probabili. Però, se le conclusioni a cui perviene possono risultare avventate, le basi da cui parte sono piuttosto solide: risultati concreti. Si tratta di due processi di direzione esattamente contraria, come risulterà chiaro nella tabella seguente, in cui cerchiamo di riassumere in uno schema quanto sinora espresso¹⁹⁸.

¹⁹⁷ Umberto Eco, op.cit., p.185.

¹⁹⁸ Potrà apparire quanto meno contraddittorio l'utilizzo di categorie della logica formale tradizionale, dopo che in precedenza si è difesa l'importanza di un approccio più legato al senso comune. In effetti, Toulmin ha definito il sillogismo "ambiguo" perché il concetto di premessa universale non permette di distinguere la garanzia dell'inferenza dal fondamento che dà autorità alla garanzia, anzi talvolta finisce per confondere tutto nei dati di partenza. Ci sono però dei casi limite in cui questa distinzione tra garanzie e fondamenti non è rilevante, ed è il caso delle argomentazioni definite da Toulmin "analitiche". Un'argomentazione è analitica se il fondamento della garanzia che l'autorizza include, implicitamente o esplicitamente, l'informazione comunicata nella conclusione stessa. Per essere certi dell'analiticità di una argomentazione, occorre effettuare il cosiddetto test della verifica: se il controllo del fondamento su cui ci si basa in una argomentazione deve *ipso facto* implicare il controllo della verità della conclusione, allora l'argomentazione è analitica. Il 'nostro' sillogismo, nel primo caso, ci sembra appunto

	ESTRAZIONE DI TIPO DEDUTTIVO		ESTRAZIONE DI TIPO INDUTTIVO
Regola	Tutte le frasi di quest'area sono rilevanti (= Quest'area è rilevante ¹⁹⁹)	Caso	Questa frase si trova in quest'area
Caso	Questa frase si trova in quest'area	Risultato	Questa frase è rilevante
Risultato	Questa frase è rilevante	Regola	Tutte le frasi di quest'area sono rilevanti(= Quest'area è rilevante)

Tabella 4: processo di estrazione testuale di tipo deduttivo e induttivo

Ancora una volta ci pare di poter affermare che siano stati i diversi retroterra culturali tra MOD1 e MOD2 a causare ragionevolmente queste differenze di comportamento: MOD1 non conosceva abbastanza i contenuti testuali da poter affermare, dopo una lettura neanche tanto approfondita, che un'area fosse particolarmente rilevante. Ciò è invece stato possibile a MOD2, che ha sfruttato al massimo il suo sapere precedente, applicando il tipo di ragionamento più adatto in quelle circostanze. MOD1 ha preferito, invece, leggere più in profondità il testo alla ricerca di una frase 'scatenante' (*trigger*), e su questa certezza (che poi una certezza non è, ma una scelta, e come tale discutibile) ha costruito il suo procedimento.

Insomma, le differenze tra 'ricercatori di informazione' producono diverse modalità di ricerca: lo affermiamo non solo sulla scorta dell'analisi fatta, che è statisticamente troppo limitata per permettere di generalizzare, ma sulla base dell'esperienza acquisita e del buon senso derivatone. Chi produce informazione ha il dovere di tener conto di queste differenze; il ricercatore universitario utilizzerà un manuale diversamente da un inesperto curioso, ma

un'argomentazione analitica, nel secondo un'argomentazione sostanziale pertanto che non può essere giudicata se non in conformità al campo in cui è espressa e al contesto in cui si attua. Se si vuole rivolgere una critica a questo schema, si può tuttavia obiettare sarebbe stato più opportuno scrivere: quasi tutte le frasi di quest'area sono rilevanti, questa frase si trova in quest'area, dunque, probabilmente, questa frase è rilevante, e viceversa. Ad ogni modo, dal momento che una tale complicazione del modello non avrebbe apportato valore aggiunto alla nostra descrizione, essa ci è sembrata poco opportuna: la semplificazione è talvolta accettabile, a patto, direbbe Toulmin, di essere consapevoli dei pericoli che comporta.

¹⁹⁹ Se consideriamo un'area di testo come l'insieme delle frasi che la compongono, trascurando eventuali complicazioni di tipo olistico, l'equivalenza fra le due proposizioni è evidente.

chi scrive il manuale dovrà sforzarsi di soddisfare entrambi.

In precedenza si è detto che i tre curatori hanno inteso semplificare il lavoro di ricerca del lettore, il quale, anziché trovarsi di fronte a quattro corposi volumi, si pone alla lettura di un manualetto che lo rimanda poi alle informazioni interessanti. Ebbene, si nota così come l'attività dei tre curatori non sia stato altro che la costruzione di una sovrastruttura testuale attraverso la quale favorire il reperimento di informazioni dal testo complessivo. Potrebbe essere paragonata, prese le debite distanze, al *software* di ricerca che accompagna di solito un *database*, e ne facilita la fruizione. Questa attività ha avuto anche un notevole valore didattico: l'efficacia sperimentale della nostra esercitazione, ancora prima del suo valore editoriale, si rivela nel fatto che confrontandoci sulle modalità con cui noi avevamo effettuato la ricerca, abbiamo potuto riflettere su come poi i lettori del testo avrebbero potuto 'completare', seguendo i loro interessi, ciò che noi avevamo cominciato. Il compito dell'autore di testi di manualistica, allora, non è quello di indirizzare il lettore in una direzione piuttosto che in un'altra, come per esempio potrebbe accadere in un saggio o in un romanzo, ma permettere a qualsiasi direzione di ricerca di risolversi in maniera efficace.

4.4. L'impostazione grafica

Una volta realizzato il processo di estrazione testuale, che ci ha permesso di selezionare una serie di brani, i quali ricostituiti in sequenza ci hanno dato un testo compresso, il problema che si è posto è stato quello di rendere fruibile questo testo.

Si trattava infatti di un testo finale di non facile lettura, sia per i 'tagli' notevoli che erano stati apportati, sia perché si presentava piuttosto disomogeneo per stile, essendo il risultato dell'affiancamento di testi scritti per tipologie di pubblico diverse e con stili diversi. Inoltre era un testo che, per quanto potesse considerarsi finito e indipendente, non poteva trascurare o nascondere la presenza dei quattro volumi d'origine. Dovevamo trovare un 'link', se così si può dire, tra il testo compresso e quello originario, che in un secondo momento si è deciso si realizzare in formato digitale²⁰⁰. Proprio da queste necessità, e cioè di favorire la lettura, di realizzare un facile

²⁰⁰ L'operazione di digitalizzazione di questi quattro testi, e la loro resa su CD-ROM, è stata un'attività in un certo senso a noi 'estranea', in quanto realizzata da un gruppo che ha operato realizzando scelte in assoluta indipendenza rispetto ai curatori del testo cartaceo. Ne faremo un rapido accenno più avanti, senza soffermarci a lungo proprio perchè si tratta di un prodotto sul cui processo di realizzazione non siamo intervenuti direttamente.

reperimento dell'informazione, di collegare libro e CD-ROM, è nata l'idea di visualizzare il testo, con l'ausilio di alcune icone, in maniera tale da rendere perfettamente visibile la sua struttura 'compressa' e le fonti d'origine.

Queste icone rappresentavano con una grafica stilizzata le copertine dei quattro testi da cui era estratta ogni frase: in questo modo, chiunque avesse aperto, anche solo casualmente, il nostro testo, avrebbe intuito che aveva di fronte un testo un po' particolare. Le frasi estratte, oltre ad essere anticipate sulla sinistra dall'icona del testo originario corrispondente, si presentano piuttosto distanziate, nella pagina, una dall'altra, per cui non c'è rischio che il lettore possa attribuire una continuità che non appartiene al testo. Al termine di ognuna di esse, un numero indica la pagina del testo originario da cui la frase è stata estratta.

L'utilizzo di queste icone prende spunto, anche se in maniera non proprio ortodossa, dalla metodologia CPP-TRS, e in particolare dagli strumenti testuali definiti segnali. Si è detto infatti che "Signs are global control icons for explaining the communicative type and function of a text unit."²⁰¹ I segnali "rappresentano per il lettore un'indicazione utile al fine di porsi in certe posizioni e di crearsi certe aspettative rispetto al testo"²⁰²; un segnale è "uno stimolo visivo che anticipa e fa scattare una azione o reazione comunicativa specifica"²⁰³. Le icone del nostro testo ricoprono perfettamente questo ruolo di anticipazione pre-testuale e stimolo visivo.

Di fatto, il lettore, di fronte a questa icona, sa che si trova di fronte ad una frase estratta, decontestualizzata una prima volta e ricontestualizzata in un nuovo ambiente comunicativo. È preparato pertanto a disporsi alla lettura in maniera adeguata, perché è chiamato a ricostruire un discorso che è stato in qualche maniera 'spezzato' e che può essere ricomposto soltanto con il suo aiuto. Inoltre il lettore si abitua rapidamente a cogliere le differenze fra le frasi presentate: breve e concise quelle provenienti da *Abitare il testo*, più discorsive quelle di *Scrittura multimediale* e *Partitura solfeggio movimento*, coinvolgenti e stilisticamente innovative quelle di *La comunicazione aziendale come arte visiva*. Sarà quindi preparato allo stile di scrittura, al tono; potrà, in base a questa consapevolezza, decidere di trascurare volutamente le frasi estratte da un testo che non lo convince particolarmente, per concentrarsi sulle altre. In ogni caso, potrà avere

²⁰¹ **Graziella Tonfoni**, *Communication Patterns and Textual Forms*, Exeter, Intellect, 1996, p.84.

²⁰² **Graziella Tonfoni**, *Abitare il testo*, Treviso, Pagus, 1995, p.63.

²⁰³ **Graziella Tonfoni**, *Abitare il testo*, Treviso, Pagus, 1995, p.17.

rapidamente un quadro sui contenuti di una tela testuale o di un oggetto, e decidere se è il caso di approfondire la conoscenza direttamente su CD-ROM (o sui testi originari, se dispone della versione cartacea), dal momento che dispone sempre della pagina da cui il testo proviene. E potrà più facilmente recuperare, in seguito, quel paragrafo o quella sezione che lo avevano interessato.

Il sistema adottato potrebbe trovare applicazione in vari tipi di manualistica tecnica. Si pensi ai manuali per l'utilizzo di un elettrodomestico, per esempio del videoregistratore. Di solito si mostrano ripetitivi per gli utenti più esperti; astrusi, incomprensibili e troppo sintetici per quelli che lo sono meno. Un tale manuale potrebbe presentare una versione 'compressa', sufficiente per l'utente più esperto; e rimandare alla versione 'ridondante' per ulteriori spiegazioni, esempi ed approfondimenti tutti gli altri che ne avessero necessità.

Le possibilità offerte poi dalle nuove tecnologie sono straordinarie: anzi, è proprio a chi si sente 'disperso' in un oceano di informazione che la compressione testuale può portare aiuto. Abbiamo già detto dell'utilità degli *abstract*, che per esempio nei CD-ROM che contengono i testi giuridici sintetizzano i contenuti della sentenza e permettono al ricercatore di informazione di comprendere immediatamente se ciò che ha di fronte è pertinente ai fini della sua ricerca. A parte il fatto che tali riassunti sono spesso, di fatto, dei testi compressi con l'aggiunta di un paio di congiunzioni e la sostituzione di qualche sostantivo; in ogni caso, se al loro posto fossero riportati degli stralci del testo fondamentali dotati dell'indicazione (numero di pagina, o di riga, o entrambi) della loro provenienza, forse il reperimento dell'informazione sarebbe ancora più immediato di quanto non lo sia adesso. Per non parlare di tutti i testi che non riportano nemmeno l'*abstract*.

Per i siti Internet la necessità di ridurre gli spazi e favorire il reperimento informativo è, come si è detto in precedenza, di importanza vitale. In rete sarebbe possibile mostrare su una sola schermata i brani estratti di un testo, invitando gli interessati a collegarsi alle pagine complete, più 'pesanti', ovvero che richiedono maggior tempo per la visualizzazione, oppure scaricarsele direttamente sul proprio disco rigido per rivederle in seguito con più calma.

Confrontiamo allora quanto da noi proposto con il parere illustre di

Umberto Eco, che sostiene in un'intervista²⁰⁴:

Nel caso di Internet, invece, solo una parte dell'informazione è di qualità. Chiunque vi abbia accesso può creare una sua homepage e metterci dentro quello che vuole. Alcuni indirizzi saranno interessanti, ma altri conterranno vera e propria spazzatura. Allora il problema è: come separare gli scarti dalla polpa? In edicola si distinguono subito i quotidiani dai settimanali, i giornali sportivi da quelli di pettegolezzo. È stata sviluppata una grafica ben precisa. Chi entra in libreria può riconoscere immediatamente un romanzo da un saggio e guardando alla serietà dell'editore può decidere se il libro è da leggere o no. Tutto ciò è impossibile con l'informazione che passa Internet. Per fare un esempio: ho trovato due indirizzi, alla voce filosofia, sulle Pagine Gialle di Internet; in pratica poi gli indirizzi sono almeno trenta, ma va bene lo stesso, anche le Pagine Gialle possono catalogare quello che vogliono. Allora, due indirizzi: uno sembra serio, l'altro è senza dubbio pieno di roba di scarto. Io, che sono del mestiere, posso distinguere tra i veri filosofi e gli eccentrici. Ma una persona qualsiasi potrebbe discernere tra queste due fonti? E io stesso, come me la caverei in un campo non mio? Nel prossimo futuro dovremo dunque impegnarci a insegnare alla gente la necessità di fare filtro, perché, se non impareranno a distinguere, a selezionare, l'accesso a tutta questa informazione sarà completamente inutile.

Nel prosieguo della conversazione Eco affermerà che se è vero che l'ormai insostenibile quantità di informazione in rete è causata dalla sua forma 'anarchica', è anche vero che non è eliminando quest'ultima, imponendo per esempio un controllo a monte dell'informazione, che si risolve il problema. Al contrario, se Internet non fosse anarchica, permettendo a chiunque di accedervi nelle vesti di produttore, perderebbe la sua fecondità, trasformandosi in un altro dei canali gestito dai gruppi dirigenti. Una rete

²⁰⁴ Umberto Eco, "Informazione telematica, verità o caos?", a cura di Paul Claessone Kevin Billingham, <http://www.fub.it/telema/TELEMA4/Eco4.html>

fatta di 'sottoreti', che, come antichi appezzamenti feudali, permettono l'ingresso solo a persone e testi selezionati, dunque, non sarebbe la soluzione del problema. La soluzione del problema, se c'è, suggerisce Eco, è di carattere formativo: bisogna insegnare a selezionare, perché solo una selezione completamente in mano al soggetto ne garantisce la libertà. La ricerca degli strumenti per operare questa selezione, venendo a mancare alcuni importanti elementi paratestuali, potrebbe richiedere parecchi anni di studio, a questo bisogna essere preparati, e far perdere inevitabilmente molto tempo.

Riguardo all'uso di strumenti elettronici 'intelligenti', Eco è piuttosto scettico: quello che bisogna sviluppare è l'intelligenza degli utenti, bisogna far crescere in loro una specie di fiuto. Molto interessante è poi una sua risposta riguardo all'inevitabile domanda sulla sopravvivenza del testo cartaceo: esso continuerà ad avere il suo spazio, come d'altronde la nostra esperienza di questi anni ha già ampiamente dimostrato. Ma questo discorso non vale per tutte le tipologie di testo. Sostiene Eco:

Ci sono due tipi di libri, quelli da consultare e quelli da leggere. I primi, probabilmente, scompariranno perché sarà più comodo tenerli sul disco. Gli altri rimarranno, perché la lettura è una attività di tipo diverso ed è più facile leggere un tascabile che lo schermo d'un computer. Consultazione e lettura sono due cose completamente differenti, l'una non esclude l'altra proprio come nulla vieta di usare la macchina pur se si sa andare in bicicletta.

Dunque ci sono testi che, essendo per loro natura pensati più per la consultazione che per la lettura, si adattano perfettamente ad essere elaborati dalla tecnologia informatica, specie se questa continuerà a progredire fornendo accessi al testo rapidi, comodi e poco costosi. Fin qui l'intervista rilasciata da Eco, sembra confermare quanto da noi affermato nel corso di questo lavoro. Non sarebbe tuttavia onesto da parte nostra nascondere la dichiarata avversità del semiotico piemontese nei confronti di alcuni strumenti di sintesi:

In campo scientifico, l'abstract, la sintesi concettuale, è uno strumento basilare. Ma io sono

contrario. Se ricevo una lettera di una rivista che mi chiede: «La preghiamo di fornirci un abstract del suo articolo», io mi rifiuto. Lo studioso, esperto della materia, è l'unica persona in grado di leggerlo e decidere se è interessante. Mentre, se ti accontenti di abstract, la funzione di filtro è stata già assolta da qualcun altro, e così magari perdi una grossa opportunità.

Eco ricorda come spesso le migliori scoperte avvengono leggendo di un libro una nota o un riferimento che parlano di tutt'altro rispetto all'argomento fondamentale del libro, e che pertanto nessun filtro avrebbe ritenuto rilevante. Il fascino dell'attività di lettura sta proprio nel regalarci 'perle' di sapere proprio nei luoghi e nei modi in cui meno ci saremmo aspettati di trovarle. Riflettendo su queste affermazioni, ci sembra tuttavia che esse non contraddicano più di tanto le tesi che noi abbiamo cercato di difendere. Per due ragioni; la prima è che Eco parla di *abstract* di suoi articoli, che sono saggi scientifici di natura filosofica e pertanto, come abbiamo avuto modo di ribadire, non rientrano nell'ambito dei testi che si adattano ad essere sintetizzati. La seconda è che egli sembra denunciare, e non si potrebbe dargli torto, un mondo accademico dove ci si 'accontenta' degli *abstract* e si trascurano i testi completi, laddove gli strumenti di compressione che noi stiamo suggerendo in nessun caso sostituiscono completamente il testo originario, ma semmai ne agevolano la consultazione. Il ruolo del selezionatore, dell'*abridger*, può essere prezioso solo nella misura in cui non è spento lo spirito critico del lettore che viene messo comunque in condizione di confrontare le scelte del selezionatore con le proprie. È chiaro che un tale confronto richiederebbe un ulteriore dispendio di tempo; ma d'altronde, ci pare chiaramente che non ci sia via d'uscita a questo bivio: da una parte il risparmio di tempo che richiede la fiducia nel selezionatore, dall'altra la perdita (ma è davvero una perdita di tempo?) se si preferisce affinare il proprio spirito critico. Infine riportiamo una frase dell'intervista con la quale potremmo 'sigillare' il nostro lavoro: "L'uso appropriato dell'informazione richiede la conoscenza dei meccanismi che l'hanno prodotta". Eco si riferisce ai meccanismi informatici, ma credo che una volta tanto sia consentita una generalizzazione anche per i processi cognitivi che abbiamo cercato di spiegare.

4.6. *Le scelte paratestuali*

In questa sezione cercheremo di mettere in luce alcuni dettagli riguardanti le scelte editoriali realizzate che, pur non essendo direttamente rilevanti per una ricerca sull'*information retrieval*, completano l'analisi dell'esperienza che abbiamo sin qui tracciato. Nella terminologia utilizzata faremo riferimento al già menzionato lavoro di Gerard Genette "Soglie. I dintorni del testo"²⁰⁵.

Genette definisce il paratesto una "soglia", una frontiera che separa il testo vero e proprio dal mondo esterno, luogo non solo di transizione ma anche di transazione. È attraverso il paratesto che l'autore o l'editore di un libro può meglio sviluppare la strategia pragmatica sul pubblico, volta ad indirizzare il lettore verso una lettura più pertinente dell'opera.

Gli elementi paratestuali di cui ci occuperemo sono il formato, il titolo, la copertina, la quarta di copertina, la prefazione e l'introduzione. Come abbiamo già avuto modo di anticipare parlando del nostro progetto, il formato, cioè il numero di pagine complessive²⁰⁶ e la loro dimensione erano già predeterminati dal fatto che il libro si sarebbe inserito in una collana.

Riguardo al titolo, scrive Genette che se il destinatario di un testo è il lettore, il destinatario di un titolo è il pubblico in senso lato, comprendente cioè non solo chi il libro lo legge, ma anche chi si limita a sfogliarlo, a discuterne, a farne un "oggetto di circolazione". La funzione del titolo è quella di identificare l'opera, designarne il contenuto e valorizzarla: di questi tre elementi, la cui influenza è variabile, solo il primo è obbligatorio. Il nostro principale obiettivo era quello di trovare un titolo che designasse i contenuti dell'opera fornendo inoltre una suggestione al lettore, accattivante quanto basta da spingerlo a sfogliare il testo o a leggere la quarta di copertina. Il risultato è andato oltre il nostro progetto, perché il titolo non solo designa il contenuto dell'opera, ma in un certo qual modo ne anticipa anche la forma espressiva. Genette parla di titoli tematici, i quali appunto designano i contenuti di un testo, o meglio i temi fondamentali ("Guerra e Pace", "Gerusalemme Liberata", eccetera) e di titoli rematici, i quali invece focalizzano l'attenzione sugli aspetti formali dell'opera ("Sonetti", "Racconti", "Massime", eccetera). Pur essendo "Il design della scrittura

²⁰⁵ Gérard Genette, *Soglie. I dintorni del testo*, Milano, Einaudi, 1989.

²⁰⁶ In realtà non si trattava di un numero preciso e rigido di pagine, ma piuttosto di un intervallo tra un numero minimo di pagine minimo e un numero di pagine massimo nel quale ci si poteva muovere con una certa elasticità.

multimediale” un titolo altamente tematico, esso conserva una sfumatura rematica, perché l’opera stessa è prima di tutto un lavoro di design di scrittura.

A questo titolo si è arrivati attraverso una collaborazione ‘creativa’ tra l’autrice e i tre curatori. Il concetto di base da cui si partiva era, ovviamente, il design del testo. “Design” è un termine che, non avendo trovato corrispondenti adatti in italiano, si è ormai diffuso nella sua dizione inglese, e può dirsi ormai assorbito dalla nostra cultura. Un termine straniero nel titolo può talvolta apparire ostico, non immediato, eccessivamente tecnico, ma nel caso di “design” questo non ci è sembrato il caso. Al contrario, da una parte questo termine connota l’idea di progettazione e pianificazione architettonica che ben si adatta ai principi della metodologia di Tonfoni, dall’altra esso richiama evidentemente *l’Information Design*, e il collegamento si mostra assai opportuno. Dunque il risultato di questa prima fase di confronti sul titolo aveva portato a “Il design del testo”. Ma tuttavia qualcosa mancava, ed era un’idea fondamentale sia per i contenuti del libro, sia per le forme espressive in cui esso aveva trovato attuazione: la multimedialità. L’aggettivo multimediale da una parte restringeva il campo semantico troppo vasto di testo, dall’altra aggiungeva una valenza notevole al nostro lavoro, che nella pluralità dei codici mediatici trova uno dei suoi capisaldi. Il passaggio da “Il design del testo multimediale” al “design della scrittura multimediale” è stato il passo successivo, e definitivo: la parola scrittura, oltre a richiamare direttamente uno dei lavori originari che sono stati compresi (*Scrittura multimediale*, appunto), focalizza l’attenzione sul processo di creazione ed elaborazione del testo, e pertanto la sua introduzione era opportuna. Ma poiché il “Design della scrittura del testo multimediale” cominciava a diventare un po’ troppo lungo come titolo, senza considerare il fatto che nella lingua italiana due complementi di specificazione consecutivi introdotti dalla preposizione “di” hanno un effetto piuttosto sgradevole, si è deciso di sacrificare il termine “testo”. La scelta è stata infine confrontata con tutte le altre diverse alternative a nostra disposizione, ed il confronto ha confermato la nostra preferenza. Ecco di seguito tutte le alternative prese in considerazione:

1. Il design del testo
2. Il design del testo multimediale
3. Il design della scrittura del testo multimediale

4. La scrittura del testo multimediale

5. Il design della scrittura multimediale

Prima di procedere, sarà bene soffermarsi su un aspetto tutt'altro che secondario: si è fatto cenno sinora alla possibilità che un testo sia sfogliato, prima di essere acquistato. Questa non è una strategia volta a vendere più copie, dal momento che in un testo scientifico, come il lavoro in questione, gli intenti commerciali sono secondari, ma un atteggiamento onesto nei confronti del lettore che non è costretto a comprare "a scatola chiusa". Ebbene, nel caso di un testo accompagnato da un CD-ROM, è praticamente inevitabile per l'editore utilizzare il cellophane per tenere insieme i due prodotti. In fase di realizzazione del testo il problema è stato posto, perché oltretutto il CD-ROM copre inevitabilmente parte della copertina, o della quarta di copertina, vanificandone in ogni caso gli intenti comunicativi. Ci si trovava in una di quelle situazioni in cui gli sforzi di un *designer* vengono vanificati da alcune costrizioni commerciali; fortunatamente la maggior parte delle librerie ha risolto il problema esponendo il libro senza il CD-ROM che poteva essere ritirato alla cassa. La potenzialità semantica della copertina pertanto veniva preservata.

L'importanza della copertina è tutt'altro che secondaria: essa è la prima manifestazione del libro che si offre alla percezione del lettore, e raddoppia la funzione che classicamente era affidata al frontespizio: quella cioè di indicare l'autore dell'opera e il titolo, oltre a fornire altre eventuali informazioni sull'editore, la collana o la data di pubblicazione. Ma limitarsi a queste informazioni vuol dire appunto ripetere quanto il frontespizio riporta, producendo informazione ridondante. Le enormi possibilità della computergrafica, tuttavia, permettono di ottenere un surplus informativo, che fa della copertina il suo veicolo. Per ciò che concerne il nostro libro, la copertina è stata realizzata da uno dei curatori, mentre gli altri due si occupavano di prefazione e introduzione: tuttavia ognuno di questi elementi paratestuali è stato oggetto di confronto. Poiché c'era accordo sulla necessità che la copertina preannunciasse, in qualche maniera, i contenuti dell'opera, non è stato difficile arrivare alla decisione di visualizzare l'idea di compressione testuale attraverso un *puzzle*, i cui tasselli provenissero dai quattro testi originali. Più in particolare, la copertina ripropone le quattro copertine dei testi originali connesse fra loro come appunto i tasselli di un *puzzle*, lasciando sullo sfondo l'immagine sfocata di un testo con alcune sottolineature, alcune cancellature, alcune correzioni: cioè le fasi embrionali

del testo organizzato che poi è stato effettivamente offerto ai lettori.

Per quanto concerne infine quarta la di copertina, la prefazione e l'introduzione, occorre fare un discorso d'insieme. Utilizzando la bella similitudine di Genette che parla del paratesto come di un vestibolo che si affaccia sul testo ma che testo ancora non è, potremmo riferirci al nostro lavoro dicendo che, nell'ordine, la quarta di copertina, la prefazione e l'introduzione sono anticamere che permettono di avvicinarsi 'gradualmente' al testo sino ad oltrepassarne la soglia.

La quarta di copertina potrebbe essere definita l'evoluzione moderna del *prière d'insérer*, l'inserto allegato al libro che forniva indicazioni sull'opera rivolte prevalentemente alla critica, o meglio, ai caporedattori delle riviste che sapevano così a quale critica specializzata sottoporre il lavoro per una eventuale recensione. Oggi la quarta di copertina ha sostituito questo allegato economicamente svantaggioso, ma soprattutto ha modificato il suo destinatario, che non è più il ristretto ambito degli addetti ai lavori, ma il pubblico più genericamente inteso. La quarta di copertina è l'elemento paratestuale che si rivolge al maggior numero di persone, e la sua importanza è pertanto fondamentale, perché deve essere in grado di fornire una succinta ma efficace descrizione dell'opera. In poche frasi, con uno stile alla portata del maggior numero di persone possibile (a meno che non si voglia deliberatamente limitare il pubblico ad un settore di specialisti, ma questo è un altro discorso), bisogna sintetizzare il contenuto dell'opera, il suo destinatario ideale, e fornire alcune informazioni sul destinatario dell'opera, presentandolo così al pubblico dei lettori. L'istanza prefattiva è già presente nella quarta di copertina, ma è nella prefazione vera e propria che si fa un passo avanti verso il testo.

La prefazione del nostro libro non si occupa ancora dei contenuti del testo, ma delle modalità espressive secondo le quali è stato progettato. Un testo particolare, innovativo, inconsueto, 'difficile' da leggere, se vogliamo, nella sua struttura compressa, deve essere presentato adeguatamente al lettore, per evitare di disorientarlo. È quello che ci siamo proposti di fare nella prefazione, in cui si sono presentate "l'origine dell'opera, le circostanze della sua redazione, le tappe della sua genesi"²⁰⁷. Si sono inoltre dichiarate le intenzioni dell'autrice e dei curatori, e di conseguenza si è reso manifesto il pubblico al quale il testo era particolarmente indirizzato, senza entrare nel merito dei contenuti della metodologia esposta. Questo ulteriore passo in

²⁰⁷ Gérard Genette, *Soglie. I dintorni del testo*, Milano, Einaudi, 1989, p.207

avanti si realizzerà più avanti, nell'introduzione. L'obiettivo della prefazione è preparare il lettore non tanto al contenuto del libro quanto alla sua espressione: egli deve essere accuratamente informato sulle scelte grafiche prese, sui motivi che hanno portato all'utilizzo del CD-ROM, sulle tecniche di compressione testuale adottate.

Nell'introduzione, finalmente, si viene a creare lo spazio per avvicinare il lettore alla metodologia CPP-TRS, alle sue potenzialità, alle sue finalità, al cammino di ricerca e di studi che hanno caratterizzato la sua evoluzione, in una parola, ai contenuti del libro. Effettuato anche questo passo, al lettore non resta che superare la faticosa soglia ed affrontare il testo vero e proprio nel migliore dei modi.

4.7. Il CD-ROM che accompagna il testo

Al CD-ROM che accompagna il testo non dedicheremo molto spazio per la semplice ragione che la sua progettazione ed implementazione tecnica è stata sviluppata da un gruppo di operatori indipendente rispetto a quello che ha curato il volume cartaceo, come segnalato nella copertina del CD-ROM stesso

Anche il CD-ROM fa uso di una serie di elementi paratestuali, quali la copertina e la postfazione. La copertina riporta le facciate dei quattro testi originari, ma stavolta non c'è elaborazione grafica né valore semantico aggiunto attraverso l'idea del *puzzle*: le quattro copertine sono riportate una accanto all'altra, in maniera piuttosto semplice, in linea con lo stile del CD-ROM, insieme al titolo dell'opera e al nome dell'autrice. Le altre informazioni sul prodotto non sono consultabili direttamente, ma sono reperibili in una postfazione accessibile durante la consultazione del CD-ROM.

Per quanto concerne la scelta dell'utente modello del CD-ROM valga quello che si è già detto del libro: si tratta di un lettore curioso, interessato alla metodologia, non necessariamente un esperto dell'area linguistico-cognitiva. Perché il prodotto finale raggiungesse questa fetta piuttosto ampia di pubblico, diveniva necessario un costo complessivo che non superasse le trenta, quarantamila lire. Questi limiti di budget hanno pesato non poco sul supporto elettronico. La decisione presa dai realizzatori per contenere i costi è stata quella di scandire le pagine dei quattro volumi originari, e riproporle in formato JPEG in due versioni di dimensioni diverse. In questo modo non si è resa necessaria la programmazione di un sistema interno di navigazione,

ma si è resa possibile la fruizione del testo attraverso un comune *browser* HTML in grado di visualizzare le immagini; al tempo stesso, non si utilizzava un sistema di lettura proprietario di qualche società, più efficace ma più costoso, come ad esempio la gamma dei prodotti Acrobat della Adobe. Il CD-ROM non presenta pertanto collegamenti ipertestuali, ma propone una lettura lineare dei volumi e non un valore aggiunto interattivo che suggerisca percorsi di lettura alternativi.

È qui doverosa una precisazione che verrà ripresa nel capitolo conclusivo, quando prenderemo in considerazione i riscontri di alcuni lettori del libro per verificare l'efficacia e gli eventuali limiti del prodotto. Tale precisazione riguarda la funzione del CD-ROM, che già nel progetto complessivo iniziale, e non solo negli sviluppi realizzati, era di natura conservativa: conservare cioè in poco spazio e con una spesa non eccessiva i contenuti di quattro volumi, per permettere il recupero dell'informazione. L'uso del CD-ROM si prestava ottimamente anche per la rapidità e la comodità di consultazione, perché in pochi istanti e con ingombro quasi nullo si raggiunge la pagina desiderata individuata dall'indicazione sul volume cartaceo. Non c'era assolutamente, da parte dei realizzatori, l'intenzione di fornire un surplus comunicativo attraverso questo supporto: un'eventuale versione ipertestuale avrebbe richiesto un lavoro di ripensamento e riscrittura complessiva dei volumi, poiché l'ipertesto non è, ci pare di aver ribadito più volte, soltanto un insieme di documenti connessi da qualche *link*, ma qualcosa di più strutturato. Fatte queste debite premesse, si potranno criticare le scelte grafiche un po' spartane dei realizzatori del CD-ROM, così come si potrà criticare la scelta di scandire le pagine trasformandole in immagini, che di fatto impedisce al lettore di intervenire sul testo con note o appunti; si potrà restare scettici sul formato grafico utilizzato, che non permette una buona resa né in fase di lettura né in fase di stampa, così come sarà possibile non essere entusiasti della assenza di mappe di navigazione. Ma tutto andrà ricondotto alla funzione principale del CD-ROM, che come si è detto è conservativa, e non accrescitiva, rispetto ai contenuti dei volumi originari, e all'obiettivo di tenere bassissimo il costo per l'utente finale. Obiettivamente, non credo proprio che si potesse chiedere di più.

5. Valutazione del lavoro

LEGGE DI CLARKE SULLE IDEE
RIVOLUZIONARIE:

Ogni idea rivoluzionaria -in campo scientifico, politico, artistico o altro- provoca tre stadi di reazione, riassumibili nelle seguenti frasi:

-È impossibile, non farmi perdere tempo;

-È possibile, ma non val la pena di farlo;

-L'ho sempre detto io che era un'ottima idea.

5.1. Un progetto messo alla prova severamente

Avviandoci alla conclusione del nostro lavoro, ci pare opportuno lasciare spazio al confronto, seppur limitato, con voci diverse rispetto alla nostra. Uno dei maggiori vantaggi, infatti, che si può trarre da una realizzazione editoriale vera e propria rispetto ad una esercitazione didattica universitaria consiste nella possibilità di testare l'efficacia dell'operazione attraverso le valutazioni di un pubblico più ampio.

Nel nostro caso, rifletteremo sulle potenzialità del lavoro di compressione e valuteremo l'efficacia di tale progetto, attraverso la verifica delle opinioni e delle esperienze di un gruppo di studenti di Milano che hanno utilizzato il volumetto come ausilio di studio nell'ambito di un progetto di insegnamento supportato a distanza attivato fra l'Università di Bologna e la Facoltà di Architettura del Politecnico di Milano e hanno messo per iscritto le loro valutazioni.

Per quanto concerne il confronto con i lavori degli studenti di Milano, è opportuno chiarire il contesto in cui questi testi sono stati prodotti e le modalità di produzione. Gli autori sono studenti universitari del corso di laurea in Disegno Industriale di Milano, e la disciplina in questione è Intelligenza Artificiale, di cui è docente il professor Marco Somalvico. Gli studenti seguono questa disciplina durante il terzo anno di lezioni, per cui non sono matricole ma studenti che hanno già raggiunto una certa competenza nell'ambito del corso di studi. Alcune delle lezioni dell'anno

accademico 1999-2000 del corso di Intelligenza Artificiale sono state caratterizzate da un seminario sulla scrittura multimediale tenuto dalla professoressa Tonfoni. Il seminario si è sviluppato in quattro ore introduttive di lezione, in cui ai ragazzi sono stati esposti i contenuti base del saggio “Intelligenza Artificiale: memorizzazione e lettura”²⁰⁸, ed è stata descritta l’attività che ha portato alla produzione de “Il design della scrittura multimediale”. In seguito i ragazzi sono stati divisi in gruppi di lavoro, costituiti mediamente da quattro o cinque studenti, cui è stata proposta una attività consistente in due fasi. La prima fase riguardava un contatto con i quattro volumi cartacei, non particolarmente approfondito, volto soprattutto ad osservarne le caratteristiche formali prima ancora che sostanziali. In seguito, nella seconda fase, ad ogni gruppo è stata assegnata una sezione di tre o quattro pagine del volume “Il design della scrittura multimediale” da leggere, con l’eventuale ausilio del CD-ROM, senza trascurare prefazione, introduzione e postfazione (quest’ultima presente sul CD-ROM). Ad ogni studente è stato chiesto di immedesimarsi in due tipi di utente: l’utente A, che necessita di acquisire rapidamente le informazioni essenziali, che predilige pertanto un testo sintetico ed immediato; l’utente B, che invece predilige una lettura che presupponga maggior livello di dettaglio nella comprensione, un lettore disposto pertanto a realizzare ulteriori approfondimenti a discapito del tempo impiegato. Rispetto a questa tipologia fornita, agli studenti è stato chiesto di compilare un diario del loro processo di lettura, e di confrontarsi con gli altri studenti del gruppo, producendo così sia un diario personale per ciascun componente sia una relazione collettiva riguardante l’intero gruppo. La scrittura del diario doveva basarsi sull’operazione di autoriconoscimento delle proprie attività cognitive, resa di fatto possibile dagli strumenti che erano stati forniti loro durante le lezioni introduttive. Ovviamente il chiarimento di ogni dubbio sulla attività proposta e l’eventuale approfondimento di alcuni temi suggeriti durante le lezioni è stato reso possibile attraverso una serie di colloqui che la professoressa ha tenuto con i vari gruppi di studenti in più fasi e dell’assistenza di Michele Semenza.

Il corpus da cui abbiamo pertanto avuto modo di attingere per verificare l’attività svolta è di conseguenza formato da 41 diari personali e da 10 relazioni collettive. Esso si è rivelato molto utile sia perché ha fornito un valido test del lavoro realizzato, sia per gli spunti per eventuali sviluppi

²⁰⁸ **Graziella Tonfoni**, *Sistemi cognitivi complessi*, Treviso, Pagus, 1991, p. 59-85.

futuri che si sono potuti trarre dalle riflessioni dei ragazzi.

Prima di addentrarci nella lettura dei risultati di questa verifica, occorre contestualizzare con dovizia l'ambiente nel quale i diari sono stati prodotti. Innanzitutto si tratta di studenti universitari particolarmente coinvolti, dal momento che sapevano che i loro lavori sarebbero stati valutati ai fini del conseguimento del voto d'esame. Il seminario ha oltre tutto rappresentato una parte di 'applicazione pratica' di un corso teorico di più ampio respiro, ed ha pertanto interessato moltissimo gli studenti che spesso nell'università italiana non trovano molto spazio per la concretizzazione di quanto appreso. Questo interesse e questo coinvolgimento, di per sé ottimali per la riuscita di un buon lavoro, hanno tuttavia generato, in qualche occasione, alcuni 'effetti perversi' di diversa natura nell'elaborazione delle riflessioni. Da una parte, c'è stato l'elogio appassionato del testo, stimolato forse dal desiderio inconscio di attrarre le simpatie del docente; dall'altra, c'è stata la critica esasperata di chi si sente dotato di un bagaglio di conoscenze superiori e vuole rendere consapevoli gli altri di questa sua superiorità.

Un altro punto da tenere presente nell'analisi di queste valutazioni è il retroterra culturale degli studenti, che, ricordiamolo, frequentano un corso di Disegno Industriale, e dunque li ha portati a concentrarsi, com'era ovvio, sugli aspetti grafici della realizzazione del testo ed in particolare del CD-ROM più che sugli aspetti cognitivi e linguistici che hanno permesso la realizzazione del prodotto. Date queste premesse, abbiamo comunque trovato estremamente stimolanti le riflessioni degli studenti, per cui abbiamo deciso di conservare le relazioni di gruppo ed una selezione dei loro diari in appendice a questa tesi.

Dal momento che, vista la particolarità del *panel* di lettori e il loro numero esiguo, un'indagine di tipo quantitativo sarebbe quanto meno azzardata, trascureremo ogni possibile approccio statistico, anche perché gli elaborati non sono stati prodotti in maniera uniforme: alcuni studenti si sono limitati al diario di lettura, altri hanno approfondito il testo con una serie di considerazioni, altri hanno realizzato una vera e propria recensione. Il nostro sarà piuttosto un approccio qualitativo: innanzitutto cercheremo di individuare quali modalità di lettura sono state preferite durante l'esercitazione; poi raccoglieremo le osservazioni, partendo dalle difficoltà e dalle critiche rivolte all'operazione, immaginando possibili orizzonti futuri, e cercando conferma alle nostre ipotesi nelle considerazioni di chi ha invece apprezzato il prodotto.

5.2. Le modalità di lettura

Abbiamo già detto dei tipi di utente in cui è stato chiesto agli studenti di identificarsi. Sarà adesso necessario riportare le modalità di lettura comunemente riconosciute e praticate, identificate dal saggio di Tonfoni²⁰⁹ che gli studenti hanno letto come ausilio alla loro attività di autoriflessione. Le presentiamo sinteticamente in questa tabella:

Modalità di lettura	Caratteristiche
Scanning	Lettura esplorativa: il lettore individua alcuni elementi informativi predeterminati (per esempio nei titoli) privilegiando una comprensione direzionata dalle scelte dell'autore.
Search reading	Lettura di ricerca: il lettore individua informazioni sulla scorta di scelte che ha prederminato autonomamente, privilegiando così l'autodirezionamento.
Skimming	Lettura globalmente informativa: il lettore ispeziona il testo sia sul piano globale che su quello locale, optando per una modalità eterodirezionata o autodirezionata, in relazione a ciò che comporta l'esplorazione testuale.
Receptive reading	Lettura ricettiva: il lettore memorizza e riproduce l'informazione selezionata. La lettura è sequenziale, lineare, non tralascia nulla e si lascia guidare dallo schema di direzionamento della comprensione prodotto dall'autore.
Responsive reading	Lettura interlocutoria: il lettore interpreta il testo come suggerimento e stimolo al proprio pensiero. La lettura, inizialmente sequenziale e lineare, presenta alcuni ritorni indietro, privilegia il ruolo attivo del lettore piuttosto che la struttura testuale.

Tabella 5: principali modalità di lettura e caratteristiche delle stesse

A queste modalità di lettura ne andrebbero aggiunte altre due, che vengono quotidianamente praticate dai lettori in quanto il loro obiettivo consiste nel minimizzare i tempi di lettura in base ad un processo di preselezione: la modalità di lettura “a macchia” e la modalità di lettura “a

²⁰⁹ **Graziella Tonfoni**, “Intelligenza Artificiale: memorizzazione e lettura”, in *Sistemi cognitivi complessi*, Treviso, Pagus, 1991, p. 59-85.

balzo e risalita”. Nel primo caso l’attenzione del lettore si sofferma su singoli elementi del testo, per allargarsi brevemente all’intorno e passare ad altro, saltando le parti intermedie del testo. Nel secondo caso, il procedimento di lettura non può provocare processi effettivi di comprensione, ma solo anticipare un approccio successivo più approfondito: il lettore “balza” dall’attacco, al centro, al finale del testo, “risalendo” poi in direzione contraria, limitandosi ad indentificare paragrafi e titoli. Questa era dunque la griglia a cui gli studenti hanno fatto riferimento per analizzare i propri processi di lettura: vediamo con quali risultati, tralasciando per il momento le difficoltà emerse, che troveranno spazio più avanti.

La maggior parte degli studenti che ha preferito identificarsi con l’utente di tipo A ha dato per scontato che questo genere di lettore non utilizzerà il CD-ROM, e che invece farà ampiamente riferimento alle immagini. Si tratta di una posizione forte che non ci sentiamo di condividere: è senz’altro vero che un utente che vuole assorbire velocemente informazioni dedicherà meno tempo ed attenzione al testo di quanto non farà chi invece vuole approfondire l’argomento; tuttavia, nulla impedisce al lettore curioso di dare un’occhiata al CD-ROM. Anche sulla funzionalità del codice visivo avremmo alcune perplessità che esplicheremo al momento opportuno. Emerge comunque da queste osservazioni una certa impostazione piuttosto rigida, schematica da parte degli studenti, quasi ‘ideologica’ nei confronti delle nuove tecnologie: esse devono rappresentare sempre e comunque un valore aggiunto di informazione, un *surplus* concettuale: l’idea che invece possano servire ‘semplicemente’ (ma non è un’operazione semplice affatto) alla gestione dei saperi tradizionalmente intesi risulta per questi fuori luogo. C’è da chiedersi fino a che punto queste posizioni siano dovute all’impostazione del corso di studi e quanto dipendano invece da un sentire culturale più diffuso.

Gli studenti che si identificano nel lettore A, dunque, non utilizzano molto il CD-ROM, ed assumono nei confronti del testo un atteggiamento di tipo esplorativo che spesso si identifica nel passaggio da una fase iniziale di modalità di lettura *scanning* ad una fase più avanzata di *skimming*; il lettore si muove soprattutto a livello locale, interessandosi alla porzione centrale del testo, ed eventualmente rallentando solo in occasione di espansioni ai lati. Giustamente gli studenti immaginano che l’utente A operi una lettura fortemente etero-direzionata, e pertanto proceda in maniera quasi lineare. Il suo modo di procedere opera passaggi avanti e indietro nel testo, tutte le volte che sorge qualche difficoltà interpretativa, e talvolta il grado di

interazione con il testo è molto alto: alcuni studenti aggiungono, al già ricco ‘*mark-up*’ del testo, una serie di annotazioni personalizzate, di rimandi interni, di rinvii ad altre sezioni. Sia chiaro che stiamo operando un notevole sforzo di generalizzazione per arrivare a tali informazioni, perché, come gli studenti stessi si sono preoccupati di sottolineare, le modalità di lettura si sono rivelate assolutamente personali e diverse a seconda dell’individuo, delle conoscenze pregresse, dell’interesse nei confronti dei contenuti. Un aspetto interessante, che potrà essere fonte di riflessioni per operazioni future di compressione testuale, è dato dalla necessità individuata da alcuni studenti di rintracciare ‘parole chiave’ suggerite dal testo, accompagnate da una serie di collegamenti e rinvii interni²¹⁰. Questa richiesta sorprende solo parzialmente: in essa si potrebbero riscontrare i primi effetti sulla nostra cultura della diffusione del *web* così com’è e così come lo abbiamo descritto nei capitoli precedenti. Un giovane che abbia dimestichezza con le nuove tecnologie è abituato alla ricerca di parole, di termini singolari, attraverso i sistemi telematici, e potrebbe aspettarsi altrettanto da un prodotto *off-line*. Di fronte a questo bisogno si possono assumere due atteggiamenti. Da una parte, si può sostenere che effettivamente la presenza di un comodo dizionarietto, sul CD-ROM, contenente le definizioni dei termini meno comuni e più tecnici può essere una buona idea per agevolare il lettore; così come una serie di suggerimenti di approfondimento realizzati attraverso richiami interni (vedi pagina...vedi il capitolo...eccetera) potrebbe agevolarne l’apprendimento. Dall’altra, però, si potrebbe replicare che è necessario porre un argine a questa sorta di ‘deriva espressiva’ che caratterizza il momento culturale che stiamo vivendo: ricercare sempre e soltanto parole isolate, infatti, dimenticando che il loro valore e contenuto è dato anche dal contesto in cui si trovano, è un modo per assecondare questo appiattimento culturale sul valore espressivo dei segni, su ‘ciò che si vede sul monitor’, o si legge sulla pagina, tralasciando l’enorme bagaglio di senso che non è immediatamente riconducibile ad una occorrenza espressiva. In quest’ottica, la ‘parola chiave’ di cui alcuni studenti sentono la mancanza, è un punto di arrivo, e non di partenza. È solo dopo aver posto l’attenzione sul testo, averne focalizzato la struttura ed aver selezionato le parti rilevanti, che il lettore può memorizzare le informazioni, riducendole per esempio a parole chiavi, connettendole fra loro in una rete di nessi logici. Questo processo deve essere svolto dal lettore; lo scrittore può agevolarlo, ma non forzarlo.

²¹⁰ L’impressione è che dietro questa necessità di parole chiavi, collegamenti e rimandi interni ci sia l’abitudine alla ricerca ipertestuale.

L'idea che ci pare sia emersa da alcune relazioni è che sarebbe opportuno presentare nel testo una serie di 'ancore', attraverso le quali il lettore segue dei percorsi predefiniti. Il problema è che in questo modo il lettore probabilmente raggiunge l'informazione, e forse lo fa in fretta, ma non la *comprende*. Per comprendere ci vuole un po' più di tempo. Sarebbe davvero un grosso guaio se la diffusione dell'ipertestualità e la sacralizzazione del concetto di velocità, tipiche di una certa cultura 'telematica' significassero, in fin dei conti, la perdita da parte del lettore della capacità di instaurare autonomamente dei nessi tra le parti del discorso e l'esigenza sempre e comunque di collegamenti imposti unicamente dallo scrittore.

Veniamo adesso al lettore di tipo B.

La modellizzazione del processo di lettura di questo utente che possiamo trarre dalla generalizzazione dei risultati delle esercitazioni è la seguente: lettura autodirezionata, di ricerca dunque, selettivo-interpretativa, con il lettore che ricerca i concetti rilevanti tenendo presente la conclusione finale del testo. Tale operazione è realizzata secondo procedimenti molto diversi a seconda del lettore, che quasi sempre porta ad una successiva fase di lettura di tipo responsivo-interlocutorio, ovvero di *responsive reading*, durante la quale al processo di lettura vero e proprio vengono affiancati momenti di riflessione relativi alle informazioni deducibili dal testo.

Vediamo adesso attraverso quali passaggi concreti gli studenti sono riusciti portare a termine il loro processo di lettura, finora analizzato solo sul piano astratto, quali difficoltà hanno incontrato e a quali fattori sono attribuiti tali difficoltà.

5.3. Le difficoltà nella lettura del testo

Gli studenti si sono concentrati, come era stato chiesto loro, sulla messa in forma del libro, più che sui contenuti della metodologia. E non sono mancate le critiche, che pare doveroso riportare.

Le critiche maggiori sono state rivolte dagli studenti che si sono identificati con il modello di utente di tipo A. Per costoro, infatti, la lettura del testo è risultata talmente ostica da divenire in alcuni momenti impraticabile. Viceversa, il testo compresso e l'interazione con il CD-ROM è risultata molto più appagante per chi cercava di approfondire autonomamente le conoscenze sulla metodologia. Elenchiamo qui di seguito le difficoltà riscontrate, anticipando tra parentesi e in corsivo alcune possibili risposte a queste obiezioni, che verranno approfondite in seguito:

1. Assenza di collegamenti interni, parole chiave, rimandi
(appesantirebbero un testo che è già piuttosto complesso);
2. Assenza di connessione logica fra le porzioni di testo estratte
(si tratta di impressioni che variano a seconda delle caratteristiche del lettore);
3. Arbitrarietà della selezione effettuata
(ogni selezione è necessariamente arbitraria);
4. Assenza di collegamenti verbali fra le stesse
(se ci fossero stati questi interventi, sarebbe cambiata la natura del testo);
5. Assenza di rielaborazione
(tale assenza è stata una scelta precisa per preservare il più possibile i testi originari) ;
6. Squilibrio nella disposizione grafica della pagina;
7. Difficoltà di comprensione delle immagini
(un'immagine non è necessariamente la spiegazione di un testo scritto);
8. Impossibilità di intervento sul testo elettronico
(anche questa è stata una scelta volta a preservare l'integrità dell'originale);
9. Difficoltà di utilizzo del CD-ROM.

Come già anticipato, il CD-ROM è stato il bersaglio preferito delle critiche rivolte dagli studenti, e più avanti vedremo perché; adesso passiamo in rassegna le altre difficoltà.

Una premessa generale è valida da qui in poi: gli elaborati hanno ampiamente confermato l'estrema importanza che in un testo come questo ricoprono la prefazione e l'introduzione (senza dimenticare la postfazione su CD-ROM). Gli studenti che hanno letto con cura queste parti sono risultati estremamente avvantaggiati rispetto agli altri. Chi ha preferito tuffarsi direttamente nel testo si è sentito spaesato, ha faticosamente cercato di comprendere autonomamente il significato di icone e figure, ha quasi sempre frainteso la struttura del testo, ed in ogni caso, anche nelle migliori delle ipotesi, ha compiuto un lavoro interpretativo aggiuntivo che gli ha solo fatto perdere tempo. Purtroppo la maggior parte dei lettori che si aggirano per le librerie si comporta nella stessa maniera: danno una sfogliata rapida, 'assaggiando' brani di testo qua e là, e se convinti da questa prima fase procedono. Se in futuro i manuali esplicativi compressi si diffonderanno, rendendone comune la fruizione, probabilmente queste prefazioni non saranno più necessarie, e il lettore che sfoglia il libro capirà subito di che si

tratta: fino a tale momento, chi non accederà al testo attraverso l'anticamera prefattiva, sarà destinato inevitabilmente a perdersi.

Passiamo in rassegna adesso le difficoltà riscontrate dagli studenti: sull'eventuale presenza di parole chiavi suggerite dallo scrittore abbiamo già detto; qui aggiungiamo che effettivamente dei rimandi interni potrebbero agevolare il lettore, ma di sicuro appesantirebbero il *mark-up* di un testo che è già di per sé notevole. Inoltre in questo modo si ridurrebbe la possibilità di movimento del lettore; si favorirebbe insomma un utente di tipo A a discapito di un utente di tipo B. Dal momento che tra gli elaborati è possibile scorgere alla stessa maniera sia richieste di maggiore intervento dello scrittore sia maggiore autonomia lasciata al lettore, ne deduciamo che tali aspetti sono fortemente personali; non potendo scrivere un testo diverso per ogni diverso lettore, bisogna mirare ad un lettore modellizzato ed accontentarsi di soddisfare quello.

La seconda difficoltà è maggiormente rilevante: il rischio più grave che corre chi realizzi un'operazione di compressione è quello di far smarrire il lettore che perde i nessi logici fra le parti di testo. Se ci fosse stata una critica unanime a riguardo da parte degli studenti, i curatori del testo avrebbero dovuto interrogarsi seriamente sulla qualità dell'operazione realizzata. Ciò tuttavia non è avvenuto: come è possibile leggere in appendice, nel gruppo 3 c'è stato chi ha trovato difficile la lettura per l'assenza di questi nessi logici, e chi invece li ha trovati chiari e comprensibili. Poiché gli studenti hanno letto le stesse pagine del testo, hanno la stessa età e più o meno la stessa preparazione culturale, è evidente che per quanto lo scrittore si sforzi di 'seminare' gli stessi contenuti nel suo testo, in alcuni casi essi saranno recepiti, in altri no, indipendentemente dall'impegno profuso a priori nell'identificare un lettore modello.

Il punto 3 ripropone la stessa difficoltà; ma in questo caso c'è un fraintendimento fra chi legge il testo e chi glielo propone. La risposta dei curatori ad un'accusa di arbitrarietà non potrà che ammettere che le selezioni sono arbitrarie. Per quanto si utilizzino dei criteri quantitativi e qualitativi riconosciuti, la valutazione della rilevanza di un brano spetterà comunque al curatore: l'operazione di compressione implica inevitabilmente da parte del curatore, una precedente operazione di interpretazione dei testi, che non può che essere personale e quindi soggetta all'accusa di arbitrarietà. È per questo che si offrono al lettore i testi completi su CD-ROM: per non sottrargli la possibilità di confrontare, se non si ritiene soddisfatto, le scelte dei curatori

con quelle che avrebbe fatto lui stesso.

Questa critica di arbitrarietà rivolta dallo studente può essere interpretata osservandola da due diverse angolature: come critica alla compressione realizzata, che non risponde evidentemente a quanto lo studente avrebbe fatto; e in questo caso l'unica risposta possibile è che la compressione proposta è una sorta di visita guidata all'interno del testo; se si preferisce percorrere i libri in un'altra direzione, si è liberi di farlo. La seconda interpretazione possibile è che la critica non è rivolta alla compressione realizzata, ma all'idea di compressione testuale in sé, che non soddisferebbe le esigenze del lettore.

Questo è il caso (e così dicendo passiamo ai punti 4 e 5) di chi lamenta l'assenza di collegamenti verbali tra le parti, che avrebbero favorito l'apprendimento. Come abbiamo già detto, questo tipo di operazione è stata brillantemente realizzata qualche anno fa da Richardson²¹¹. Non c'è dubbio che un'eventuale rielaborazione del testo avrebbe reso più scorrevole la lettura anche nel nostro caso. Il suggerimento esplicito di qualche studente è stato quello di inserire delle frasi 'cerniera', scritte dai curatori, per legare i vari brani estratti. L'idea è realizzabile: semplicemente, però, in questo caso, non ci troveremmo più di fronte ad un testo compresso, ma ad un particolare tipo di riassunto. Per ottenere un risultato accettabile quanto a scorrevolezza ed immediatezza non sarebbe sufficiente infatti aggiungere dei connettivi tra un brano e l'altro (quindi...inoltre...in conseguenza di ciò...) ma sarebbe necessaria una seppur parziale riscrittura dei brani stessi, se non altro per rispettare le concordanze. Se si considera poi il diverso stile con cui i libri originari sono stati scritti, si comprende quanto difficoltoso sarebbe stato un intervento che da una parte agevolasse la lettura e dall'altro rispettasse la sintassi originaria dei brani. Il testo compresso, con la sua indiscutibile 'sgradevolezza' nella lettura lineare sequenziale, con il suo nascondere alcune parti per evidenziarne altre, è uno stimolo alla lettura più accurata. Suggerisce al lettore alcuni elementi basilari, invitandolo all'eventuale approfondimento. In nessun caso vuole presentarsi come esauriente. Ricordiamo ancora una volta che esso è uno strumento per il reperimento dell'informazione, un 'accessorio' al testo, più che un testo completo in sé. Sforzarsi di completarlo, di nascondere i suoi limiti, 'camuffarlo da riassunto' significherebbe tradirne la natura.

²¹¹ **Graziella Tonfoni**, *Writing As a Visual Art*, Intellect, Exeter, 1994 (con J.E. Richardson revisore tecnico, e una prefazione di Marvin Minsky; seconda edizione 2000).

Non paia che le critiche rivolte non siano state prese in considerazione con la dovuta attenzione. Al contrario, dalla lettura di questi elaborati si coglie un rilevante suggerimento per il futuro: un utente di tipo B probabilmente si trova maggiormente a suo agio rispetto ad un utente di tipo A, che si sente spaesato di fronte al testo compresso. Per quest'ultimo, forse, sarebbe più efficace una guida alla metodologia lineare, tradizionale. Sarebbe una buona idea per il futuro se non fosse che questa guida, per quanto riguarda la metodologia CPP-TRS, esiste già anche in italiano²¹². Valga comunque come suggerimento in linea generale per la teoria dell'*Information Retrieval*.

Veniamo adesso al secondo 'polo' di discussione estrapolabile dai testi degli studenti: l'utilizzo della grafica e la difficile comprensione delle immagini. Premesso che sicuramente la competenza a riguardo di chi studia Disegno Industriale è maggiore rispetto alla nostra, si accettano di buon grado i suggerimenti di chi consiglierebbe un maggiore equilibrio della pagina, con più spazio per i testi e meno per le icone (anche se poi, in direzione completamente opposta, vanno le affermazioni di chi lamenta una scarsa visibilità delle icone). Ad onor del vero, i commenti riguardanti l'impostazione grafica sono stati per la maggior parte di apprezzamento; c'è un punto, però, già accennato in precedenza, che merita qualche ulteriore valutazione.

Praticamente tutti gli studenti, con un'analogia di vedute che evidentemente deve avere radici profonde, ritengono che il ruolo delle immagini sia quello di spiegare il testo. Si ritengono soddisfatti quando questa funzione esplicativa è svolta ottimamente dalla figura, permettendo loro di sperimentare quello che il testo dice; molto meno quando avvertono qualche dissonanza che li costringe ad un ulteriore sforzo di astrazione e interpretazione. Il ruolo gerarchicamente 'inferiore' del linguaggio visivo rispetto a quello verbale scritto, assorbito dalla cultura esterna, è confermato dall'idea diffusa per cui le immagini servono soprattutto agli utenti A, mentre sono accessori inutili per chi ha intenzione di approfondire la conoscenza. Ebbene, questa sorprendente idea di subordinazione dei codici, anziché di integrazione effettiva degli stessi, è stata quella che ci ha stupito di più, e con la quale siamo in completo disaccordo. L'immagine non si limita a 'spiegare' i contenuti testuali; altrimenti non si spiegherebbe la sua posizione nella pagina, che anticipa il testo, anziché seguirlo. L'immagine veicola un

²¹² Ci riferiamo al testo a cura di **Laura Vignoletti**, "Guida alla conoscenza e all'apprendimento della metodologia CPP-TRS di Graziella Tonfoni. Mappa di navigazione a cura di Laura Vignoletti", in *Quaderni di Ricerca Linguistica*, Serie Strumenti, Parma, Edizioni Zara, 1993.

contenuto che il testo non sarà mai in grado di fornire nella stessa maniera, attraverso gli stessi stimoli e producendo le stesse evocazioni; suggerisce accostamenti visivi, figure metaforiche, procedimenti cognitivi di comprensione che semmai il testo potrà confermare o meno. Multimediale, allora, non è ripetizione dello stesso contenuto attraverso più codici, ma integrazione di contenuti diversi attraverso codici diversi che portano poi alla definizione di un oggetto. Chi si occupa di arti visive da sempre 'multimediali', come il cinema, sa perfettamente che un'immagine è in grado di veicolare informazioni che un testo scritto trasferisce a fatica e con minor resa.²¹³

Questo procedimento, applicato al testo esplicativo, non sempre facilita l'immediatezza della lettura, al contrario; chi osservi una figura concepita secondo tali modalità, e legga in seguito un testo che fa riferimento allo stesso oggetto, potrà non comprendere subito il senso, essere costretto a fare un gioco di sponda tra testo e immagine, accantonare momentaneamente la comprensione per ritornarci successivamente. Ma è solo attraverso questi 'spostamenti' avanti e indietro nel testo, che non dovrebbero generare frustrazione ma semmai stimolare l'interesse, che il lettore arriva gradualmente a padroneggiare un contenuto, 'ne diventa padrone'.

Gli ultimi due punti del nostro schema riassuntivo riguardano il testo in formato elettronico. Riportiamo, per dovere informativo, le critiche rivolte al CD-ROM, anche se interverremo solo quando esse coinvolgeranno direttamente l'argomento principale della nostra discussione, cioè il reperimento dell'informazione.

La prima critica ha riguardato la portabilità del testo elettronico, che non è compatibile con i sistemi Macintosh: problema che sarà senz'altro risolto in futuro, anche perché la scelta di un formato graficamente 'povero' come quello dell'HTML accompagnato da figure JPEG trova le sue ragioni soprattutto nella portabilità che non è legata a particolari *software* proprietari. Numerose critiche sono state rivolte a questa scelta, che offre una visualizzazione delle immagini meno chiara.

Altre critiche da parte di studenti sono state rivolte al sistema di navigazione interno al CD-ROM, alla sua povertà grafica, all'impossibilità di intervenire sul testo. Queste note saranno sicuramente preziose per gli sviluppatori del CD-ROM, ma non possono che riguardarci marginalmente.

²¹³ Si legga a riguardo **Vincenzo Cerami**, *Consigli a un giovane scrittore*, Torino, Einaudi, 1996.

Più interessante dal nostro punto di vista è una critica generale, già ampiamente preannunciata, rivolta al testo elettronico, accusato di non sfruttare le potenzialità del supporto dal momento che non presenta elementi ipertestuali. Abbiamo già detto abbastanza sulla questione, sull'impossibilità di 'tradurre' un testo lineare in un testo ipertestuale senza apportare interventi di riscrittura. Oltretutto, i testi di Tonfoni che sono stati compressi contenevano già un alto grado di innovatività nelle tecniche di scrittura che avrebbero solo complicato, se non vanificato, una resa ipertestuale. Ciò che merita una riflessione è però la compattezza delle opinioni degli studenti, che di certo non possono essere considerati un campione rappresentativo di tutto pubblico di lettori ma che sono comunque altamente qualificati a dare giudizi, riguardo all'associazione immediata tra testo elettronico e ipertesto. C'è dunque una cultura molto radicata che associa il supporto testuale, la materialità su cui si poggia il testo, con le tecniche di scrittura che lo realizzeranno. Non è una novità assoluta: anche nell'universo 'cartaceo', infatti, chiunque si aspetterebbe di leggere un quotidiano su ampi fogli ruvidi e non rilegati, un fumetto d'autore su fogli patinati abbastanza grandi, con copertina rigida e accurata rilegatura, e così via. Si instaura cioè un'abitudine nel lettore che crea un nesso tra sostanza dell'espressione e contenuti. Tuttavia, per quanto riguarda la carta stampata queste associazioni sono il risultato di decenni, se non di più, di consuetudini consolidate; il testo elettronico su CD-ROM non ha più di quindici anni, e già crea aspettative altrettanto specifiche.

Restiamo tuttavia convinti del fatto che tale associazione possa essere smentita, e che il testo lineare possa comunque trovare spazio su CD, sfruttandone i bassi costi, l'ampia gamma di materiale memorizzabile, l'interazione con sistemi di reperimento dell'informazione cartacei (come nel caso del nostro testo) o elettronici (i sistemi esperti o i *search-engine*). Semmai, bisognerebbe stimolare maggiormente la capacità decisionale dell'utente che si muove autonomamente tra vari supporti, se gliene viene offerta l'occasione. È preoccupante, per esempio, constatare come a nessuno degli studenti che abbia trovato difficoltà nella fruizione del CD-ROM sia venuto in mente di stampare le pagine su carta per una comoda lettura tradizionale. È un peccato, un autentico spreco, irrigidirsi su particolari abitudini di lettura, soprattutto quando esse ci impediscono di fruire del testo in maniera più soddisfacente.

Manca ancora una educazione al multimediale che renda il lettore capace

di sfruttare pienamente le tecnologie di scrittura e lettura senza vincolarsi ad un mezzo particolare; se ciò è vero per alcuni studenti universitari, si può ben immaginare come debba essere la situazione se solo si provi ad allargare il contesto fruitivo. In ogni caso, spetta per primi ai produttori di informazione 'far crescere' il proprio pubblico fornendogli strumenti efficaci di reperimento informativo ed educandolo a farne uso.

5.4. I riscontri positivi alla lettura del testo

Abbiamo volutamente dedicato molto spazio alla critiche riguardanti le difficoltà riscontrate dagli studenti durante la lettura sia perché le abbiamo ritenute più stimolanti, in un'ottica rivolta al futuro, sia per evitare di autocompiacerci improduttivamente nelle lodi e negli apprezzamenti. Tuttavia le lodi sono state numerose, e non vanno trascurate. Vediamo di presentare sinteticamente la caratteristiche del testo maggiormente apprezzate:

1. Immediata acquisizione della struttura portante della metodologia;
2. Disposizione testuale degli elementi chiara ed ordinata;
3. Capacità di stimolare il lettore curioso invitandolo ad approfondire;
4. Ampia possibilità di personalizzazione della lettura
5. Facilità di accesso all'informazione.

La struttura della pagina, con la sua insolita disposizione degli elementi testuali, la presenza 'estranea' di icone e rimandi intertestuali, l'ampio spazio lasciato alle figure, ha senz'altro colpito l'attenzione degli studenti. Ha inoltre permesso loro di inquadrare immediatamente una 'struttura portante' della metodologia, un'ossatura che si è rivelata importante e funzionale come griglia di lettura per eventuali approfondimenti. La semplicità e l'ordine delle pagine sono stati elementi determinanti, come era facile sospettare, per la facilitazione dell'apprendimento, per bilanciare, se vogliamo, le difficoltà di lettura inevitabili, già ricordate, presentate da un testo compresso.

Dopo la comprensibile sorpresa generata dal trovarsi di fronte ad un testo così poco comune, è stato testimoniato dai lettori il sorgere di un forte interesse, una viva curiosità che li ha portati ad approfondire il discorso. Talvolta anche troppo, verrebbe da dire: i ragazzi, probabilmente, sono stati talmente stimolati dal testo compresso, da sentirsi poi delusi dal CD-ROM, verso il quale nutrivano maggiori aspettative.

Tutti gli studenti, nei loro lavori, hanno sottolineato in varia misura la diversità di approccio alla lettura, ma alcuni hanno fatto un passo successivo importante, e cioè si sono resi conto che questa diversità di approccio è stata resa possibile e agevolata in ampia misura dal testo stesso. La personalizzazione del testo va intesa qui infatti come la concreta possibilità di una struttura testuale di adattarsi alle esigenze del lettore, ed è ovviamente di grado differente se si parla della guida telefonica o dei *Miserabili* di Hugo. Nel nostro caso possiamo dirci soddisfatti dell'elasticità che il testo presenta, anche se, come abbiamo già notato, è un'elasticità che ha mostrato di adattarsi meglio ad un utente di tipo B che ad un utente di tipo A. Spingersi oltre, nel tentativo di soddisfare proprio tutti i bisogni degli utenti possibili, è un po' azzardato, perché si corre il rischio di alleggerire a tal punto l'istanza autoriale da renderla sfumata ed ineffabile. Nulla impedisce tuttavia che tale sfida possa essere colta in futuro.

Nonostante tutte le difficoltà emerse, gli 'attriti' dovuti agli elementi di novità che queste tecniche comportano nelle tradizionali abitudini di lettura, così come le aspettative non sempre giustificate nei confronti del multimediale, una buona parte degli studenti ha evidenziato la facilità con la quale è stato possibile reperire le informazioni. Questa facilità è stata spesso espressa esplicitamente, talvolta è stato possibile leggerla tra le righe dei loro commenti. Se si considera che la quasi totalità degli studenti non conosceva la metodologia di Tonfoni prima del seminario, e che dopo una breve lettura del testo che non ha portato via loro molte ore mostrava di averne compreso gli elementi di base, non si può che essere soddisfatti del risultato raggiunto. Questa 'prova' così severa, visto il pubblico di lettori che hanno giudicato il testo, è risultata pertanto estremamente preziosa: da una parte le critiche ben motivate rivolte non possono che stimolare ulteriori miglioramenti in futuro; dall'altra, gli apprezzamenti e i buoni risultati conseguiti dagli studenti nelle loro 'passeggiate' testuali, confermano che la strada intrapresa è quella giusta.

Il progetto di compressione testuale realizzato, non dimentichiamolo, rappresenta soltanto il primo passo di un cammino che dimostrerà soltanto col tempo le effettive potenzialità delle tecniche di *information design* applicate al testo con lo scopo di agevolare il reperimento informativo. È un progetto che se apprezzato potrà essere imitato, approfondito, attualizzato in modi diversi e con diversi risultati. Di sicuro resta da compiere molto lavoro affinché, nella cosiddetta 'società dell'informazione', ai lettori sia concesso

concretamente il diritto di reperire autonomamente e in assoluta libertà critica le informazioni disponibili. Ed è augurandoci che tale progresso si attui pienamente in un futuro non molto lontano, che chiudiamo il nostro lavoro, speranzosi di avere, seppure in piccolissima parte, contribuito alla realizzazione di tale progresso.

Appendice: elaborati dei gruppi di studenti²¹⁴

Gruppo 1

L'analisi delle pagine 17/23 del libro "il design della scrittura multimediale" è stata effettuata separatamente: solo dopo aver analizzato le nostre personali modalità di lettura ci siamo confrontati ed abbiamo rilevato i punti in comune e le differenze nell'approccio e nella comprensione del testo.

Il primo punto da evidenziare è che entrambi ci siamo immedesimati in un **lettore di tipo B**, cioè con un approccio approfondito e analitico nei confronti del testo. Ciò ha sicuramente fatto sì che le nostre modalità di lettura siano state molto simili, sebbene con profonde differenze dal punto di vista della "simpatia" nei confronti della costruzione concettuale di questo tipo di manuale.

Un'ulteriore analogia è rappresentata dal fatto che tutti e due **abbiamo letto il testo due volte**: la prima in maniera veloce, per capire a grandi linee come fosse costruito il testo, sia dal punto di vista del contenuto che dal punto di vista grafico, la seconda in maniera molto più approfondita, cercando di comprendere in modo completo i contenuti. Analizzando attentamente le peculiarità delle modalità di lettura utilizzate, però, abbiamo notato notevoli differenze, tra le quali l'uso dello schema riassuntivo.

L'ultimo punto in comune emerso dal confronto delle nostre modalità di lettura è rappresentato dall'identificazione della **tipologia del testo**: entrambi siamo rimasti perplessi dalla sua atipicità, dal fatto che esso è costituito da stralci di altri quattro testi.

Questo fattore ci ha portato a considerazioni profondamente diverse.

Simone, infatti, ha considerato il **testo come un indice** in funzione del CD allegato, quindi non ne ha notato particolari pecche dal punto di vista della completezza

²¹⁴ Gli elaborati dei gruppi sono stati riportati integralmente senza tagli nè modifiche, se si eccettuano quelle necessarie a conformare il testo al formato di pagina di questo lavoro. Pertanto, la presenza di eventuali parole in grassetto, o di soluzioni grafiche particolari è dovuta agli studenti e non ad una evidenziazione da parte nostra.

dell'informazione; nonostante ciò, però, sottolinea come la lettura su schermo dei testi non sia particolarmente agevole.

Letizia, invece, lamenta una **incompletezza delle informazioni** ed una **arbitrarietà** delle parti scelte: il testo può essere utile per una veloce consultazione, ma non è assolutamente adatto come strumento di studio; esso viene visto come un Bignami, dove tutto viene accennato e nulla approfondito.

Inoltre il CD non viene valutato come un valido strumento di supporto, perché toglie il contatto fisico con il testo, la possibilità di sottolineare, di "vivere" le informazioni; inoltre indubbiamente la lettura a monitor è stancante e poco adatta allo studio.

Simone considera un vantaggio il fatto di avere i quattro testi compresi in un unico CD, per la **facilità di trasporto**, mentre Letizia sostiene che comunque il fatto di **necessitare di un computer** per poterli leggere sia una grave pecca.

Ulteriori considerazioni riguardano la **distribuzione grafica** del testo: secondo Simone lo spazio che separa i paragrafi è sufficiente a far capire che essi appartengono a contesti diversi, che sono dei "quanti" di informazione. La presenza delle icone, invece, crea della confusione, poiché esse sbilanciano visivamente la pagina e sembrano avere importanza riguardo al contenuto del paragrafo a cui sono affiancate, mentre sono solo dei riferimenti pratici.

Anche Letizia è disturbata dalla presenza di queste icone, poiché le ritiene troppo complesse dal punto di vista grafico, sbilanciate tra loro e nei confronti della pagina e di difficile interpretazione.

Rapporto tra il libro e il CD-ROM

Dopo l'analisi della modalità di lettura dei capitoli, abbiamo esaminato l'approccio con il CD.

Naturalmente il tipo di uso è decisamente diverso, proprio per le grandi differenze che esistono tra il supporto cartaceo e il monitor.

L'informazione contenuta sul CD è notevolmente più completa, ma ci è sembrato che la corrispondenza tra i paragrafi e le parti del CD non fosse poi così immediata. C'è infatti un richiamo alla pagina di appartenenza, ma

non alla zona precisa di testo contenente il paragrafo in esame.

Anche l'organizzazione del CD lascia a desiderare, poiché ci è sembrato che le potenzialità di questo tipo di supporto siano state sprecate: **poteva infatti essere interessante usare una struttura di tipo ipertestuale**, per integrare tra loro i quattro testi e facilitare il lavoro degli utenti nella ricerca di informazioni e nella comprensione dei contenuti.

Infine la disposizione grafica del CD è troppo legata a quella di un libro, e ci sembra essere una struttura poco adatta per la consultazione su monitor.

Gruppo 2

Il processo di autosservazione della lettura del libro della professoressa Graziella Tonfoni *Il design della scrittura multimediale* condotto in un primo momento singolarmente ci ha portato a riflessioni comuni e divergenti nel metodo di approccio (di lettura e di comprensione) al testo.

Per un migliore approccio al libro abbiamo inizialmente letto l'introduzione del *Design della scrittura multimediale* e la postfazione del CD-ROM.

Avendo adottato il punto di vista dell'utente "A", un lettore curioso che vuole conoscere i contenuti principali, noi tutte abbiamo privilegiato una lettura veloce in vista di una rilettura più approfondita delle parti precedentemente evidenziate. La seconda lettura risulta rispetto alla prima più lenta e riflessiva.

Le parti del testo evidenziate nella prima lettura sono risultate essere per tutte noi quelle di maggiore importanza o di maggiore difficoltà per la comprensione degli argomenti.

Sono state riscontrate delle differenze per quanto riguarda il riferirsi alle immagini durante la lettura: alcune vi fanno riferimento per capire meglio attraverso più codici espressivi, altre le ignorano.

Tutte noi, quando ci troviamo di fronte a termini o parti di frasi non immediatamente comprensibili, tendiamo a continuare nella lettura per raggiungere almeno una comprensione globale del testo.

Per quanto riguarda il ricorso al CD-ROM tra noi si sono riscontrate modalità differenti: alcune vi ricorrono di volta in volta quando si presenta nel testo una parte poco chiara (alla fine del paragrafo), altre soltanto quando sono giunte alla fine del capitolo. Il tipo di lettura del CD-ROM è comune a tutte noi: leggiamo rapidamente, passando da un inizio di frase a quello successivo per individuare chiarimenti e spiegazioni. La facilità di accesso e navigazione del CD-ROM è invece risultata diversa: alcune hanno trovato il metodo per raggiungere comodamente le pagine desiderate, altre si sono scontrate con il lento procedimento dello scorrimento delle pagine. Tutte abbiamo riscontrato scomoda all'interno del CD-ROM l'individuazione dei termini e delle frasi non chiare presenti nel libro: sarebbe agevole trovare richiami, segni o sottolineature nel testo di ciò che stiamo cercando senza dover scorrere l'intera pagina ipertestuale per individuarlo. Nella consultazione del CD-ROM, per chiarire il significato di alcune parti del libro, spesso ci troviamo a rileggere frasi molto simili che ripetono gli stessi concetti senza aggiungere qualcosa di nuovo alla comprensione del libro-raccolta. Più in particolare ritroviamo regolarmente all'inizio di ogni paragrafo una frase molto sintetica e poco esauriente, per cui ricorriamo alla consultazione del CD-ROM che riporta la stessa frase, senza perciò chiarirne il significato; questo è probabilmente dovuto al fatto che il libro a cui si rimanda, *Abitare il testo*, è già una sintesi dei libri precedenti.

In ultima analisi l'idea di raccogliere in un CD-ROM i quattro volumi con un libro guida di riferimento ci sembra utile e flessibile per i diversi tipi di lettori che affrontano l'argomento delle metodologie di progettazione della scrittura multimediale.

La sintesi degli argomenti presenti nel libro-raccolta ci risulta efficace per un lettore curioso che necessita di un assorbimento veloce delle nozioni principali.

Gruppo 3

Il gruppo, formato da 4 persone, ha lavorato sui paragrafi 9 e 10 ("Tela dell'interazione dialogica" e "Tela della scrittura e della lettura collettiva") del testo "Il design della struttura multimediale".

Due dei componenti del gruppo: Francesca e Liborio, si sono immedesimati nell'utente che necessita di acquisire informazioni rapidamente, quindi quel lettore che predilige un testo sintetico in cui i concetti sono espressi in modo immediato senza giri di parole; mentre Fabio e Matteo si sono identificati nel lettore che ha bisogno di successivi approfondimenti.

Al di là dei contenuti trattati che prescindono dalla nostra analisi, abbiamo notato durante la "revisione" di gruppo che i particolari comunicativi riscontrati individualmente sono ricorrenti anche se le valutazioni sono risultate alquanto diverse ed in alcuni casi addirittura opposte. Nonostante questo, bisogna comunque sottolineare come ognuno di noi ha notato una serie di particolari diversi che tutti insieme si completano tra loro formando un quadro d'insieme abbastanza omogeneo e puntuale. Questo perché, nonostante ci siamo classificati come utenti dalle diverse esigenze, ognuno di noi ha comunque delle modalità di lettura differenti e personali.

All'interno del testo analizzato sono stati riscontrati i seguenti particolari comunicativi: le immagini, i blocchi testuali e le icone di riferimento. Le rispettive valutazioni di merito sono però risultate assai varie. Ecco un elenco di considerazioni fatte da ogni singolo componente del gruppo (per più precise spiegazioni si rimanda alle singole relazioni personali):

- Liborio (utente A) ha riscontrato salti logici tra un blocco di testo e l'altro, nonché difficoltà nell'associare un concetto a una immagine, un linguaggio tecnico lontano da un utente che non conosce a priori l'argomento e una struttura testuale rigida e poco elastica nei confronti del lettore.

- Francesca (utente A) ha notato una discontinuità logica e visiva tra le diverse parti del testo, uno stile incoerente (dovuto essenzialmente al fatto che le varie frasi sono tratte da libri diversi), una scarsa scorrevolezza nella lettura, una difficile interpretazione dei vari paragrafi e difficoltà di comprensione delle immagini perché troppo astratte e perché posizionate prima del testo.

- Matteo (utente B) ha criticato il fatto che le immagini a inizio testo sono troppo astratte tanto che implicano la necessità di aumentare la concentrazione per

poter essere comprese, ma soprattutto ha trovato poco funzionale la decisione di proporre estratti di un discorso più complesso senza una certa rielaborazione, infine come utente di sistema Macintosh, ha notato una realizzazione del CD-ROM poco curata nell'aspetto della portabilità.

- Fabio (utente B) è quello che più si discosta dalle considerazioni degli altri componenti del gruppo anche se, in sintonia con le opinioni degli altri per quanto riguarda una scarsa chiarezza e una scarsa spiegazione delle immagini poste sotto il titolo e per il fatto che il CD-ROM sia poco immediato e funzionale. Per le altre cose ha notato che è sempre garantito un nesso logico tra i blocchi testuali abbastanza lineare e scorrevole, nonostante derivino da libri caratterizzati da stili diversi; oltre ad aver notato una certa facilità di reperire gli approfondimenti grazie al numero di pagina riportato tra parentesi quadre alla fine di ogni blocco testuale.

Riportate brevemente le valutazioni di ogni elemento del gruppo, possiamo dire che le conclusioni relative alla modalità di lettura di questi due paragrafi non sono univoche, ma presentano varie sfaccettature.

Per i due utenti A (Francesca e Liborio), il modo di presentare il testo non è quello ottimale, anzi, va letto e riletto più volte prima di essere capito; inoltre per una completa comprensione risulta necessaria comunque la lettura degli altri libri e questo va contro l'esigenza primaria di ottenere informazioni rapidamente.

Per quanto riguarda gli altri due componenti abbiamo opinioni differenti: per Matteo, probabilmente il libro analizzato è più utile per un lettore che già ha avuto occasione di confrontarsi con i testi principali e voglia fare il punto della situazione sugli argomenti trattati. In quest'ottica il libro così organizzato vorrebbe essere una sintesi, un riassunto di elementi base per un eventuale richiamo logico degli altri libri. Diversa è la conclusione di Fabio: per lui il testo così organizzato è ottimale per un successivo approfondimento: ciò che rende il tutto poco funzionale è il modo con cui si ricerca tale approfondimento. Il lettore, secondo Fabio, sarebbe più facilitato a cercare il relativo approfondimento direttamente sui libri piuttosto che sul CD-ROM: le pagine

sono solo digitalizzate e la ricerca della pagina di riferimento è lunga e non immediata.

Gruppo 4

L'analisi individuale del libro "Il design della scrittura multimediale" ci ha portato a delle considerazioni comuni, nonostante lievi discordanze, legate alle abitudini, alle esperienze ed alle conoscenze di ogni componente del gruppo.

Innanzitutto ciascuno di noi è stato fortemente condizionato dalla struttura imposta dall'autrice G. Tonfoni: coscienti del fatto che il testo è un'estrazione delle parti più rappresentative dei quattro precedenti volumi "Scrittura multimediale", "La comunicazione aziendale come arte visiva", "Partitura, solfeggio, movimento" e "Abitare il testo" della stessa G. Tonfoni, ognuno di noi è stato indotto ad un analogo approccio cognitivo: considerare il testo di "Il design della scrittura multimediale" non in modo sequenziale, ma come summa di più parti.

La struttura del libro ha fatto sì che al primo approccio concentrassimo la nostra attenzione sull'impostazione e distribuzione grafica del testo nella pagina; siamo stati subito portati a focalizzare gli elementi portanti della struttura:

- il testo diviso evidentemente in paragrafi
- le icone a margine che rappresentano il titolo del libro da cui è stato estrapolato il paragrafo
- le immagini in primo piano

Questo primo tipo di lettura, che siamo portati ad avere, dunque:

- è di tipo esplorativo e sintetico
- è guidata dalle scelte dell'autrice
- porta il contenuto in secondo piano
- è legata ad una modalità di comprensione direzionata (scanning).

Successivamente e parallelamente ogni componente del gruppo, per avere un'idea più approfondita del tema trattato, ha svolto una lettura narrativo-sequenziale veloce: una lettura informativa sia a livello globale che

locale del testo, individuando le parole chiave da approfondire successivamente in una lettura a macchia.

Per una migliore memorizzazione abbiamo infine riconsiderato ogni singolo paragrafo all'interno del suo contesto originale nel CD-ROM.

Alcuni di noi, pensando così di avere una visione più completa ed un più facile processo d'assimilazione, hanno inoltre appuntato a lato dei paragrafi delle considerazioni estrapolate dal CD-ROM.

Abbiamo scelto di essere dei lettori di tipo B che approfondiscono le informazioni date attraverso l'utilizzo integrato e in parallelo del CD-ROM e del testo, riuscendo ad avere così un ruolo attivo ed autodirezionato e ponendo la stessa attenzione in ogni parte del testo (e non localizzata solo sulle premesse o sulle conclusioni).

Siamo partiti da una modalità di lettura predirezionata dall'autrice per poi passare ad una lettura personale e a rimandi (interlocutoria) per procedere poi ad un'autonoma elaborazione del testo in base agli obiettivi che ci eravamo posti.

Differenza testo scritto e CD-ROM

In base alle attitudini personali alcuni di noi si sono trovati meglio nella consultazione del libro piuttosto che del CD-ROM: se il libro da una parte permette di personalizzare con sottolineature e richiami il testo che da una condizione di anonimato passa ad uno stato personalizzato ed interpretato a seconda dell'interlocutore, il CD-ROM dall'altra dà una più completa visuale dell'argomento trattato e permette una maggior facilità d'interazione.

Gruppo 5

Nella lettura del libro "Il design della scrittura multimediale" e precisamente dei capitoli inerenti la miniatura ed il polittico testuale, tre membri del gruppo si sono identificati nell'utente A, ovvero in colui che necessita di un assorbimento rapido e veloce dell'informazione, mentre il resto del gruppo, precisamente tre soggetti, si sono identificati nell'utente B, intendendo con esso un utente che desidera avere una conoscenza approfondita dell'argomento esaminato.

La discussione e la comparazione delle singole esperienze di lettura in gruppo, ci hanno permesso di delineare parametri, metodologie e situazioni a volte concordanti tra loro, a volte fortemente differenziate.

Il punto di partenza comune a tutti i membri del gruppo è dato dalla non conoscenza dell'argomento trattato che conseguentemente ha generato un processo di lettura rapido, lineare e direzionato dall'autore, volto a fornirci una conoscenza generale, di base degli argomenti.

La modalità di lettura utilizzata in questo primo momento, è riconducibile allo *scanning*, per alcuni tale metodologia si è alternata a quella dello *skimming*.

A questa prima lettura ne è seguita una seconda, tale operazione è stata eseguita da tutti i membri del gruppo ed ha seguito modalità differenti in particolare in relazione all'obiettivo perseguito.

Per alcuni di noi questa fase è stata caratterizzata esclusivamente dalla rilettura del testo cartaceo, per altri tale operazione è stata affiancata, in alcuni momenti, all'utilizzo dei testi presenti sul CD-ROM, ed in altri casi ancora la seconda lettura è stata circoscritta ai testi digitali.

Coloro che si sono identificati nell'utente A hanno fatto ricorso solo marginalmente al CD-ROM, affiancandolo alla lettura del testo cartaceo ed utilizzandolo per chiarire determinati passaggi, senza però leggere l'intero testo digitale.

L'operazione di lettura è stata impostata su un processo di *skimming*, una modalità di lettura globalmente informativa che esamina il testo sia sul piano globale, sia su quello locale, prevedendo eventuali operazioni di rilettura per chiarificazioni.

Diverso è stato l'utilizzo dei testi digitali per coloro che si sono identificati nell'utente B, in questo caso tali testi sono stati utilizzati per approfondire le informazioni acquisite nella lettura del testo cartaceo.

Le modalità di lettura utilizzate da questi utenti, in questa fase, sono differenti e vanno dalla lettura "a rilievo" seguita da un'operazione di chiosatura, riconducibile alla *Receptive Reading*, e ancora questa modalità seguita ed affiancata alla *Responsive Reading*.

Differenti sono risultate le modalità di scorrimento del testo in relazione al tipo di utente scelto ed alle modalità di lettura adottate.

Gruppo 6

In questo lavoro ci siamo divisi i compiti in modo da impersonare nella lettura utenti di tipo A (3 persone) e di tipo B (2 persone). I primi sono lettori che non hanno mai affrontato l'argomento in questione, i secondi sono invece lettori esperti, che hanno già affrontato e acquisito altrove i contenuti che si presentano nel testo scelto, e che sono semmai interessati ad approfondirlo. Rileviamo che l'utente A si avvicina al testo nelle linee più generali possibili con una fase di "attenzionamento", ovvero cercando di muoversi attraverso titoli e prefazioni per capire soprattutto quali sono gli ambiti del discorso e in che direzione l'autore muove la presentazione dei vari paragrafi. Importante anche la presenza delle icone che fanno capire il tipo di riferimento bibliografico. Quindi l'utente A si muove con una prima scansione del testo con una sorta di "receptive reading" in modalità di lettura "d", non mancando di abbinare la lettura veloce all'osservazione, anche se superficiale, ai disegni, i quali aiutano a esemplificare i contenuti letti. E' a questo punto necessaria una lettura più accurata (modalità "h"), orientandosi con le icone di rimando per seguire il discorso eventualmente anche in testi paralleli. Dall'analisi di modalità di lettura effettuate da utenti di tipo B, risulta che il testo è affrontato in prima analisi con una visualizzazione globale dei contenuti (search reading), ma questa volta servirà per capire quali ambiti del discorso sono interessanti. Segue una modalità di lettura a macchia, affrontata soprattutto in modalità "d" ed "f". L'analisi più accurata per estrapolare i cardini più interessanti avverrà poi in modalità "l", utilizzando le icone di rimando come strumento per ricercare un approfondimento (che è poi l'obiettivo centrale di una simile lettura). Da notare che è risultato evidente che l'utente di tipo B non utilizza il testo in modo lineare come presentato dall'autore, ma costruendosi una propria gerarchia di riferimenti-guida che gli servono per muoversi più liberamente in modo da effettuare un approfondimento più veloce e personalizzato. In questo senso le icone sono state utilizzate di più dall'utenza B.

I disegni invece sono stati più utili ai lettori di tipo A, in quanto per gli altri sono per lo più elementi che confermano conoscenze già acquisite. Dall'analisi sembra risultare che ciò che differenzia maggiormente un utente "di tipo B" da uno "di tipo A", sia il desiderio di superare un livello superficiale (ma non "superfluo") di comprensione del testo in favore di un'attenzione, più analitica e "matura", principalmente dovuta ad un interesse particolare nei confronti del suo contenuto.

Gruppo 7

Dall'analisi delle relazioni individuali è emerso che sono presenti alcuni elementi comuni nelle modalità di lettura. Tuttavia i singoli punti di vista presentano delle sfumature che contraddistinguono le diverse personalità e i diversi interessi.

Dalla discussione è risultato che la lettura per ogni componente del gruppo è stata affrontata con il medesimo obiettivo, in quanto si è effettuata una lettura globalmente informativa: abbiamo cioè fatto uno *Skimming*.

Il primo passo è stato quello di analizzare la struttura portante delle pagine, cercando di capirne l'impostazione. In seguito la parte assegnata è stata letta per intero nell'ordine dato dallo scrittore, seguendo la sequenza dei capitoli, lo si è quindi percorso indiscriminatamente per avere un primo approccio generale sul testo. Successivamente si è passati a rivedere e rileggere il testo per controlli e chiarificazioni approfondite attraverso la consultazione del CD-ROM, capitolo per capitolo. Ci siamo resi conto che essendo un testo a noi nuovo e non conoscendone il contenuto, era necessario applicare tale modalità di lettura. Pensiamo che questo tipo di lettura sia stata motivata anche dalla brevità del testo che non porta a dover saltare i contenuti per avere una veloce visione di insieme.

Da qui il tipo di lettura per alcuni casi si è diversificata.

In due casi la lettura dei paragrafi scritti per intero visibili su CD-ROM è stata affrontata in maniera analoga alla lettura effettuata sulle pagine del libro. Però ciò si è affievolito dopo i primi paragrafi in quanto era chiara la loro struttura. La lettura si è quindi trasformata in una più veloce e non sequenziale. Sono

stati scorsi velocemente pezzi di testo tanto da capire a grandi linee il contenuto, con lo scopo di raggiungere e soffermarsi su quei punti che spiegassero meglio ciò che sul libro era stato accennato, ma non era stato pienamente compreso. Tale tipo di lettura lo si ritrova nella definizione di *search reading*, "che consiste nella individuazione di informazioni relative a scelte predeterminate autonomamente dal lettore" (*Sistemi cognitivi complessi*, Graziella Tonfoni, Pagus edizioni).

Il tipo di scorrimento che è stato applicato è stato quello che nel testo "*Sistemi cognitivi complessi*" viene definito come modalità "L"; infatti lo scorrimento del testo è stato condotto proprio in funzione di cercare determinati elementi informativi e concetti che fossero pertinenti rispetto all'assunto di base. In questo modo la lettura è stata più veloce, meno analitica delle precedenti, definibile autodirezionata-rapida.

Negli altri casi la lettura del testo integrale presente nel CD-ROM, è stata rapida e superficiale. Si è utilizzata una modalità selettivo/interpretativa ovvero, è stato interpretato ed astratto il concetto globale all'interno dei singoli capitoli ricercando, con una lettura "a macchia", quelle parole chiavi presenti nel libro.

Ciò che è emerso dal lavoro di gruppo è il fatto di aver appreso l'esistenza e l'utilità dei diversi approcci alla lettura, che permettono ad un utente di avere la padronanza dei concetti teorici di apprendimento, e di rendere quindi, con la loro applicazione consapevole, più efficace e rapida la lettura di un testo. Inoltre, il fatto che ci sia un libro che contenga gli estratti portanti dei quattro volumi incuriosisce il lettore spingendolo a leggere il testo integrale all'interno del CD-ROM. Ciò è un efficace strumento di sintesi che permette un rapido apprendimento dei concetti fondamentali. Infine la lettura e' stata influenzata dal tipo di "compito" richiesto. Se per esempio fosse stata necessaria per memorizzarne i contenuti o per carpire informazioni utili per un'applicazione dei concetti espressi in una progettazione, essa sarebbe stata direzionata verso differenti modalità di lettura.

Gruppo 8

Il lavoro ha avuto un seguito piuttosto graduale;

inizialmente siamo partiti dalla lettura del materiale pervenutoci via e-mail dal prof. Michele Semenza, per comprendere, anche se in modo piuttosto approssimativo, quelli che erano i contenuti dei quattro libri da consultare. Successivamente abbiamo letto singolarmente le pagine assegnateci (da 96 a 102) del libro "Il design della scrittura multimediale"; L'approccio alla lettura è stato differente tra i vari membri del gruppo.

Confrontando, poi, le diverse esperienze, abbiamo riscontrato la presenza di tre persone che hanno affrontato il testo come utenti "A" e due che invece lo hanno affrontato come utenti "B".

Gli utenti A, non avendo sufficiente conoscenza dell'argomento, hanno avuto un approccio alla lettura completamente dipendente dal percorso tracciato dall'autore. Possiamo, quindi, definire questo tipo di operazione come "receptive reading"; il percorso degli utenti è stato in principio lineare, sequenziale e poco approfondito al fine di assimilare le informazioni necessarie e sufficienti a raggiungere una visione globale degli argomenti trattati dall'autore (modalità "d"). Nessun singolo elemento all'interno delle pagine (titoli, immagini, icone, paragrafi) ha avuto un ruolo predominante sugli altri. A questo primo scorrimento del testo non abbastanza esaustivo, è seguito, in ogni modo, un approfondimento operato anche con l'utilizzo del CD-ROM e quindi la modalità di lettura è passata al tipo "h" mantenendo comunque un livello di apprendimento del testo ad uno stadio superficiale.

Gli utenti "B", possedendo una discreta conoscenza della materia, hanno avuto una maggiore autonomia nel gestire le informazioni presentate. Dopo una "scansione" iniziale utile a considerare il testo nel suo insieme, hanno focalizzato la loro attenzione solo su particolari elementi. Questo tipo di lettura è definito "search reading". Da questi elementi sono passati all'adozione di un procedimento "a macchia" in cui l'analisi è stata più approfondita e determinata dalla volontà del lettore, senza seguire alcun ordine stabilito a priori. Le modalità di lettura adoperate sono state quelle di tipo "h" ed "l" secondo le quali l'utente non deve sottostare agli schemi proposti dall'autore ma è libero di agire attivamente, operando delle scelte, secondo le sue esigenze di approfondimento.

Gruppo 9

Traendo le conclusioni riguardo ai procedimenti, alle modalità ed agli approcci di lettura adottati dal gruppo, anche se l'analisi è stata effettuata su un numero ridotto di casi, è possibile giungere alla formulazione di alcune ipotesi, illustrate qui di seguito.

Gli atteggiamenti dei lettori sono stati diversi sia a livello di approccio sia a livello di modalità, e non si può certo giudicare quale sia il procedimento corretto e quale quello scorretto, è importante invece individuare i fattori che hanno portato a quelle differenze. Noi come causa scatenante abbiamo individuato gli obiettivi: il lettore infatti, come si evince dalle analisi dei singoli componenti, per prima cosa, quando deve rapportarsi con un qualsiasi testo ha sempre degli obiettivi prefissati, che possono rimanere invariati o mutare durante la lettura o la rilettura dello stesso, ma che comunque esistono.

Sono dunque gli stessi obiettivi a determinare il tempo di interazione tra lettore e testo, ed il tipo di atteggiamento inteso come modalità di lettura dell'utente rispetto all'oggetto in questione.

Gruppo 10

Prima di intraprendere la lettura del libro, abbiamo identificato due tipologie di utenti. Due di noi hanno utilizzato un approccio alla lettura di tipo A, gli altri due di tipo B.

La metodologia di lettura di tipo A identifica un utente che vuole dedicare poco tempo a quest'operazione e trarne le informazioni essenziali, all'opposto l'approccio di tipo B presuppone un maggior livello di dettaglio nella comprensione a discapito del tempo impiegato.

Nonostante l'impiego della stessa metodologia per gli utenti A (come per gli utenti B) si è riscontrato un approccio personale quindi diversificato per i quattro lettori.

Gli utenti A hanno raggiunto il loro obiettivo tramite percorsi diversi; il primo lettore ha scelto un itinerario di lettura personale legato alle icone che contrassegnano il testo, il secondo lettore invece ha selezionato dal testo le parti di suo interesse dando maggiore attenzione

ai titoli e alle immagini. Entrambi hanno attribuito molta importanza al processo di attenzione e selezione per estrapolare solo le informazioni essenziali di loro interesse.

Anche i lettori B hanno compiuto tale operazione, ma come fase iniziale di un processo più approfondito. Dopo una attenta lettura del testo, hanno analizzato gli approfondimenti contenuti nel CD-ROM.

Il primo lettore B ha riletto più volte i singoli paragrafi per una comprensione di base per poi scendere nei dettagli tramite i riferimenti ipertestuali.

Il secondo lettore ha invece assimilato tutti i concetti con una sola lettura per poi rimandare interamente all'analisi del CD-ROM.

Da quest'analisi si nota come a seconda dell'approccio dell'utente e dei suoi obiettivi, si possono trarre informazioni differenti da un medesimo testo.

Nel nostro caso, questo strumento didattico si è verificato compatibile alle diverse metodologie di lettura; la sua stessa struttura definita dal supporto cartaceo e dal formato digitale (il CD-ROM) rispecchia i vari possibili approcci. L'utente A può quindi usufruire solo del libro mentre l'utente B dedica buona parte del tempo anche alla consultazione del CD-ROM. Il testo è stato reso dai curatori in modo tale da venire incontro alle esigenze di entrambi i lettori e di permettere loro una facile fruizione delle informazioni.

Elenco bibliografico dei testi consultati²¹⁵

ALLINSON LESLEY, HAMMOND NICK

- 1999 “A learnig support environment: the hitch-hiker’s guide”, in McAleese Ray, *Hypertext. Theory into practice*. Exeter, Intellect.

ANTINUCCI FRANCESCO

- 1997 *Ipermedia e processi di apprendimento*.
[Http://www.infosys.it/INFO90/obbligo/antinucc.html](http://www.infosys.it/INFO90/obbligo/antinucc.html)

BALLMER THOMAST.

- 1980 *Frames e strutture contestuali*. Parma, Ed.Zara, 1984. Ed.or.: “Frames and Context Structures. A Study in a Procedural Context Semantics with Linguistic Applications in Sentence and Text Linguistics”, in Ballmer T., Kindt W., Bujke H. (a cura di), *Zumthema Sprache und Logik*, Hamburg, Verlag.

BARRETT EDWARD (ED.)

- 1989 *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press.

BARTHES ROLAND

- 1979 “From work to text”, in *Textual Strategies*, edited by Josué V. Harari, Cornell University Press, Ithaca.

BARZON GIULIANA, BERTINI PATRIZIA, SORDI STEFANIA

- 1999 *Trovar lavoro in editoria. Da freelance a imprenditore. Una guida completa alle professioni del libro*. Milano, Editrice Bibliografica.

²¹⁵ Questa breve bibliografia riporta dapprima le informazioni sull’opera nella versione in cui è stata consultata, ed in seguito le altre eventuali versioni esistenti. Nella maggior parte dei casi, tuttavia, si tratta di testi in inglese di cui non esiste una traduzione italiana. L’anno riportato sotto il nome si riferisce alla prima edizione.

BEAUGRANDE ROBERT ALAIN DE, DRESSLER WOLFGANG ULRICH

1981 *Introduction to Text Linguistic*, London, Longman.

BELLMAN NEROZZI PATRIZIA (A CURA DI)

1997 *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

BETTETINI GIAMPAOLO

1996 *L'audiovisivo. Dal cinema ai nuovi media*. Milano, Bompiani.

BLASI GIULIO

1997 *Dentro e fuori di Internet, osservazioni sulla nascita di un nuovo medium. Dispense per il corso di Teoria e tecnica degli ipertesti multimediali*, Università di Bologna, A.A. 1997-98

BLOOM HAROLD

1973 *The Anxiety of Influence. A Theory of Poetry*, New York: Oxford University Press. (tr.it., *L'angoscia dell'influenza*, Milano, Feltrinelli, 1983.

1994 *The Western Canon: the Books and School of the Ages*, New York, Harcourt Brace

BOLTER JAY DAVID

1991 *Lo spazio dello scrivere. Computer, ipertesti e storia della scrittura*. Milano, Vita e Pensiero, 1993. Ed.or. *Writing space- The Computer, Hypertext abd The Hystory of Writing*, Hillsdale, N.J.

BORIOSI GIACANELLI ELENA, ASCARI DIANA

1995 *Guida alle ricerche bibliografiche. Dalla biblioteca alle banche dati alle reti telematiche*. Bologna, Zanichelli.

BROWN MARLIN C.

1999 *Human-Computer Interaction Design Guidelines*, Exeter, Intellect.

BUSH VANNEVAR

1945 “As we may think”, in Nyce M. James, Kahn Paul (eds.), *From Memex to hypertext. Vennevar Bush and the Mind's Machine*, San Diego, Academic Press, 1991.
<http://www.isg.sfu.ca/~duchier/misc/vbush>

CADIOLI ALBERTO

1997 “Le soglie dell’ipertesto”, in Bellmann Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.
1998 *Il critico navigante [saggio sull’ipertesto e la critica letteraria]*, Genova, Marietti.

CARLSON PATRICIA ANN

1989 “Hypertext and Intelligent Interface for Text Retrieval”, in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press.

CAZALE BERARD CLAUDE, MORDENTI RAUL

1997 “La costituzione del testo e la «comunità degli interpreti»”, in Bellman Nerozzi Patrizia(a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

CERAMI VINCENZO

1996 *Consigli a un giovane scrittore*, Torino, Einaudi.

CIOTTI FABIO

- 1997 “Testo rappresentazione e computer. Contributi per una teoria della codificazione informatica dei testi”, in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

COLLINS RANDALL

- 1988 *Teorie Sociologiche*, Bologna, Il Mulino, 1992. Ed. or. *Theoretical Sociology*, Orlando, Harcourt Brace Jovanovich, Inc.

COLOMBO CHIARA

- 1997 “Teoria e prassi della codifica del testo nell’analisi computerizzata dell’epistolario di J.G. Hamann”, in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

CONKLIN JEFF

- 1987 “Hypertext: An Introduction and Survey”, in *IEE Computer* n° XX(1987)

CONTE MARY ELIZABETH (A CURA DI)

- 1977 *La linguistica testuale*. Milano, Feltrinelli.

COOMBS JAMES H., RENEAR ALLEN H., DEROSE STEVEN J.

- 1977 “Markup Systems and the Future of Sholarly Text processing”, in *Communications of the ACM*, V.30, n°11, Novembre.

DELLA CASA MAURIZIO

- 1994 *Scrivere Testi*, Firenze, La Nuova Italia.

ECO UMBERTO

- 1975 *Trattato di semiotica generale*, Milano, Bompiani..
1977 *Come si fa una tesi di laurea*, Milano, Bompiani.
1979 *Lector in fabula*, Milano, Bompiani.
1991 *I limiti dell'interpretazione*, Milano, Bompiani.

EDWARDS M. DEBORAH, HARDMAN LYNDA

- 1999 “ ‘Lost in Hyperspace’: cognitive mapping and navigation in a hypertext environment”, in McAleese Ray , *Hypertext. Theory into practice*. Exeter, Intellect.

GENETTE GÉRARD

- 1987 *Soglie. I dintorni del testo*. Milano, Einaudi, 1989 Ed. or. Seuil, Paris, Édition du Seuil.

GIGLIOZZI GIUSEPPE

- 1997 “Critica letteraria e nuove tecnologie”, in Bellmann Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

GREGG LEE W., STEINBERG E.R. (EDS.)

- 1980 *Cognitive Processes in Writing*, Hillsdale, New Jersey, Erlbaum.

HAUBEN MICHAEL, HAUBEN RONDA

- 1997 *Netizens. On the History and Impact of Usenr and the Internet*. Ieee Computer Society, Los Lamitos, California.

HAYES JOHN R., FLOWER LINDA S.

- 1980 "Identifying the Organization of Writing Processes", in Gregg Lee W., Steinberg E.R. (eds.), *Cognitive Processes in Writing*, Hillsdale, New Jersey, Erlbaum.
- 1980 "The Dynamics of Composing", in Gregg Lee W., Steinberg E.R. (eds.), *Cognitive Processes in Writing*, Hillsdale, New Jersey, Erlbaum

HJELMSLEV LUIS

- 1959 "L'analisi strutturale del linguaggio", in *Saggi Linguistici*, Bari-Roma, Laterza, 1981 Ed.or. *Essais linguistiques. Travaux du Cercle Linguistique de Copenhague*, Copenhague, Nordisk Sprog-og Kulturforlag.
- 1959 "Per una semantica strutturale", in *Saggi Linguistici*, Laterza, Bari-Roma, 1981. Ed.or. *Essais linguistiques (Travaux du Cercle Linguistique de Copenhague)*, Copenhague, Nordisk Sprog-og Kulturforlag.
- 1959 "Lingua e 'parole' ", in *Saggi Linguistici*, Laterza, Bari-Roma, 1981. Ed.or. *Essais linguistiques (Travaux du Cercle Linguistique de Copenhague)*, Copenhague, Nordisk Sprog-og Kulturforlag.
- 1961 *I fondamenti della teoria del linguaggio*. Torino, Einaudi, 1987. Ed.or. *Prolegomena to a Theory of Language*, Madison, University of Wisconsin.

HOLT O'BRIAN PATRICK

- 1989 *Models of Writing: A Question of Interaction?*, in Williams Noel, O'Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect.

HOVSTAD UNNI

- 1989 "Computer Assisted Essay-writing: an Interdisciplinary Development project at Eikeli Grammar School" , in Williams Noel, O'Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect.

LANDOW GEORGE P.

- 1992 *Iper testo, il futuro della scrittura*, Bologna, Baskerville, 1993. Ed.or. *Hypertext. The Convergence of Contemporary Critical Theory and Tecnology*, Baltimore, The Johns Hopkins University Press.
- 1997 “The Convergence of Contemporary Critical Theory and Tecnology”, in Bellman Nerozzi Patrizia (a cura di), *Internet e le muse. La rivoluzione digitale nella cultura umanistica*. Milano, Mimesis.

LESK MICHAEL

- 1997 *Preserving Digital Objects: Recurrent Needs and Challenges*, <http://community.bellcore.com/lesk/auspres/aus.html>

MARCHIONINI GARY

- 1995 *Information Seeking in Electronic Environments*. Cambridge, University Press.

MARLIN C. BROWN

- 1999 *Human-Computer Interaction Design Guidelines*. Exter, Intellect.

MEYROWITZ NORMAN

- 1989 “The Missing Link: Why We’re All Doing Hypertext Wrong”, in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press.
- 1991 “Hypertext- Does It Reduce The Cholesterol, Too?”, in Nyce M. James, Kahn Paul (eds.), *From Memex to hypertext. Vennevar Bush and the Mind’s Machine*, San Diego, Academic Press.

MCALEESE RAY (ED.)

- 1999 *Hypertext. Theory into practice*, Exeter, Intellect.

MINSKY MARVIN

1985 *La società della mente*, Torino, Einaudi, 1989. Ed.or. *The society of mind*, 1985.

MORRIS CHARLES

1938 *Foundations of the Theory of Signs*, Chicago, University of Chicago Press, 1959.

NELSON TED

1981 *Literary Machines 90.1.*, Padova, Muzzio Editore, 1992 (traduzione italiana dell'edizione del 1990). Prima edizione: *Literary Machines*, Swarthmore (Pa.) pubblicato in proprio.

NIELSEN JACOB

Hypertext 87. [Http://www.sun.com/950523/columns/alertbox/ht87.html](http://www.sun.com/950523/columns/alertbox/ht87.html)

NYCE M. JAMES, KAHN PAUL (EDS.)

1991 *From Memex to hypertext. Vennevar Bush and the Mind's Machine*, San Diego, Academic Press.

PETÖFI JÁNOS S.

1977 "Semantica, pragmatica, teoria del testo", in Mary Elisabeth Conte (a cura di), *La linguistica testuale*, Milano, Feltrinelli.

POTTER GARY

1997 *A Post-Modern Type of Thing*.
[Http://mbhs.bergtraum.k12.ny.us/~tnellen/hypertext.html](http://mbhs.bergtraum.k12.ny.us/~tnellen/hypertext.html)

RESSLER SANDY

1997 *The Art of Internet Publishing, The Internet and Beyond*, Upper Saddle River, Prentice Hall.

SAUSSURE FERDINAND DE

1922 *Corso di linguistica generale*, Roma-Bari, Laterza, 1992. Ed.or. *Cours de linguistique générale*, Paris, Editions Payot.

SCAVETTA DOMENICO

1992 *La metamorfosi della scrittura. Dal testo all'ipertesto*. Firenze, La Nuova Italia.

SERAFINI MARIA TERESA

1992 *Come si scrive*, Milano, Bompiani.

SHARPLES MIKE, GOODLET JAMES, PEMBERTON LYN

1989 “Developing a Writer’s Assistant”, in Williams Noel, O’Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect.

SMOLIAR W. STEPHEN, JAMES D. BAKER, TAKEHIRO NAKAYAMA, LYNN WILCOX

1997 “Multi-Media Search: An Authoring Perspective” , in Smeulders W.M.Arnold, Jain Ramesh, *Image Databases and Multi-media Search*. London, World Scientific Publishing.

SMEULDERS W.M.ARNOLD, JAIN RAMESH, (EDS.)

1997 *Image Databases and Multi-media Search*. London, World Scientific Publishing.

SMORTI ANDREA

2000 “Con il verde si passa? Il posto della violazione in un mondo di norme: il ragionamento acrobatico” in *Psicologia Contemporanea*, n°157, gennaio-febbraio 2000.

SPADARO ANTONIO

2000 “Il libro e internet: le tecnologie digitali e il mondo dell’editoria”, in *La Civiltà Cattolica*, N°3591, gennaio 2000.

2000 “Le antiche radici teologiche dei recenti dibattiti sulla lettura”, in *La Civiltà Cattolica*, n°3593, marzo 2000.

SPINA PAOLA

1997 *Parole in rete*, Firenze, La Nuova Italia.

STEFANI ANNA

1998 “Ci si può fidare della struttura HTML dei documenti Web?”, in *Web Marketing Tools*, n°12, dicembre 1998.

TANEMBAUM ANDREW S.

1988 *Reti di calcolatori*, Torino, Utet, 1997. Ed.or. *Computer Networks*, Prantice Hall, NJ, 1988.

TONFONI GRAZIELLA

1991 *Sistemi cognitivi complessi*, Treviso, Pagus.

1991 *Scrittura multimediale*, Treviso, Pagus.

1991 *La comunicazione aziendale come arte visiva*, Treviso, Pagus.

1992 *Partitura, solfeggio, movimento*, Treviso, Pagus.

- 1992 “La multimedialità attiva” in *Media Duemila*, n°4, aprile.
- 1994 *Writing As a Visual Art*, Exeter, Intellect (U.K.); Scarecrow Press, Maryland, (U.S.) con J.E. Richardson revisore tecnico, e una prefazione di Marvin Minsky; seconda edizione 2000.
- 1995 *Abitare il testo*, Treviso, Pagus.
- 1996 *Communication Patterns and Textual Forms*, Exeter, Intellect.
- 1998 *Information Design. The Knowledge Architect's Toolkit*, Scarecrow Press, Lanham, Maryland.
- 1998 “Intelligent control and monitoring of strategic documentation: a complex system for knowledge miniaturization and text iconization”, in *The Proceedings of the ISIC/CIRA/ISAS 98 Conference*, Gaithersburg, National Institute of Standards and Technology, U.S. Dept. of Commerce, MD.
- 1999 “On augmenting documentation reliability through communicative context transport” in *The Proceedings of the 1999 Symposium on Document Image Understanding Technology*, Annapolis, MD.
- 1999 *Il design della scrittura multimediale*, Napoli, Cuen, 1999.

TOULMIN STEPHEN

- 1958 *Gli usi dell'argomentazione*, Torino, Rosenberg & Sellier, 1975. Ed. or. *The use of Argument*, London, Cambridge University Press, 1958.

VALENTE M.

- 1999 *Motori di ricerca alla massima potenza*, in “Internet News” n°2 febbraio.

VIGNOLETTI LAURA

- 1993 “Guida alla conoscenza e all'apprendimento della metodologia CPP-TRS di Graziella Tonfoni. Mappa di navigazione a cura di Laura Vignoletti”, in *Quaderni di Ricerca Linguistica*, Serie Strumenti, Parma, Edizioni Zara.

WILLIAMS NOEL

- 1989 "Computer assisted writing software: Ruskin", in Williams Noel, O'Brian Holt Patrick, *Computers and Writing: models and tools*. Exeter, Intellect.

WITTGENSTEIN LUDWIG

- 1922 *Tractatus logico-philosophicus*, Torino, Einaudi, 1964. Ed.or. *Tractatus logico-philosophicus*, Londra, Kegan Paul, Trench, Trubner & Co., Ltd.
- 1953 *Ricerche filosofiche*, Torino, Einaudi, 1967. Ed.or. *Philosophical investigations*, Oxford, Basil Blackwell.

ZIMMERMAN MURIEL

- 1989 "Reconstruction of a Profession: New Roles for Writers in the Computer Industry", in Barrett Edward (ed.), *The Society of Text. Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, The MIT Press.

Indice

<i>Introduzione</i>	2
<i>1.L'Information Retrieval: dalla prospettiva cognitiva a quella computazionale</i> __	8
1.1. Il punto di vista della linguistica computazionale	8
1.2. Alcuni principi semiotici rilevanti	10
1.3. Un contributo semantico alla risoluzione di alcuni problemi computazionali __	12
1.4. La logica del ragionamento comune	15
1.5. Il ragionamento metaforico e la sua rilevanza per l' <i>information retrieval</i> _____	24
1.6. <i>L'Information Design</i>	30
1.7. La metodologia CPP-TRS	33
<i>2. Le nuove tecnologie per l'elaborazione testuale</i>	37
2.1. La scrittura con il computer	37
2.1.2. Il computer come macchina da scrivere	40
2.1.3. Il computer come ausilio intelligente alla scrittura	43
2.2. Le caratteristiche multimediali del testo elettronico: l'ipermedia _____	47
<i>3.Alcuni strumenti informatici per l'information retrieval</i>	65
3.1. Editoria on-line ed editoria <i>off-line</i>	65
3.2. Reperire informazioni <i>off-line</i>	67
3.3. Reperire informazioni in rete	74
3.4. I limiti dei motori di ricerca	83
3.5. Il <i>mark-up</i> e le possibilità offerte all' <i>information retrieval</i>	91
3.6. La 'migrazione' dell'informazione	102
3.7. Brevi cenni sulle applicazioni dei sistemi esperti per l' <i>information retrieval</i> __	106
<i>4. Dalla teoria alla realizzazione pratica: studio di un caso</i>	110
4.1. La genesi del testo compresso	110
4.2. La compressione testuale	117
4.3. Il ruolo dell' <i>abridger</i>	119

4.4. L'impostazione grafica _____	136
4.6. Le scelte paratestuali _____	142
4.7. Il CD-ROM che accompagna il testo _____	146
5. <i>Valutazione del lavoro</i> _____	148
5.1. Un progetto messo alla prova severamente _____	148
5.2. Le modalità di lettura _____	151
5.3. Le difficoltà nella lettura del testo _____	154
5.4. I riscontri positivi alla lettura del testo _____	161
<i>Appendice: elaborati dei gruppi di studenti</i> _____	164
Gruppo 1 _____	164
Gruppo 2 _____	166
Gruppo 3 _____	167
Gruppo 4 _____	170
Gruppo 5 _____	171
Gruppo 6 _____	173
Gruppo 7 _____	174
Gruppo 8 _____	175
Gruppo 9 _____	177
Gruppo 10 _____	177
<i>Elenco bibliografico dei testi consultati</i> _____	179
<i>Indice</i> _____	191

Dedicato alla memoria di Eduardo