

$$\frac{A_{10}}{265}$$



Fabio Ciotti

# Il testo e l'automa

*Saggi di teoria e critica computazionale  
dei testi letterari*



Copyright © MMVII  
ARACNE editrice S.r.l.

[www.aracneeditrice.it](http://www.aracneeditrice.it)  
[info@aracneeditrice.it](mailto:info@aracneeditrice.it)

via Raffaele Garofalo, 133 A/B  
00173 Roma  
(06) 93781065

ISBN 978-88-548-1129-4

*I diritti di traduzione, di memorizzazione elettronica,  
di riproduzione e di adattamento anche parziale,  
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie  
senza il permesso scritto dell'Editore.*

I edizione: aprile 2007

## Indice

<b>INDICE</b> .....	<b>5</b>
<b>INTRODUZIONE</b> .....	<b>7</b>
<b>CODIFICA E MODELLIZZAZIONE DEL TESTO</b> .....	<b>11</b>
CODICI E MODELLI .....	13
TESTO E MODELLI DEL TESTO .....	15
MODELLI DEL TESTO E CODIFICA DIGITALE .....	19
<b>I LINGUAGGI PER LA CODIFICA DEL TESTO</b> .....	<b>21</b>
SCHEMI E LINGUAGGI DI CODIFICA.....	23
I REQUISITI DI UNO SCHEMA DI CODIFICA PER LA RICERCA UMANISTICA .....	23
I TIPI DI SCHEMI DI CODIFICA .....	28
XML (EXTENSIBLE MARKUP LANGUAGE) .....	39
LA TEXT ENCODING INITIATIVE .....	42
<b>ASPETTI TEORICI E METATEORICI DELLA CODIFICA</b>	
<b>TESTUALE</b> .....	<b>47</b>
LA NATURA EPISTEMOLOGICA DELLA CODIFICA TESTUALE .....	48
TEORIA DEL TESTO E STRUTTURA DEI LINGUAGGI DI CODIFICA.....	52
OLTRE I LIMITI DELLA TEORIA GERARCHICA .....	60
<b>LA CODIFICA COME LINGUAGGIO TEORICO</b> .....	<b>65</b>
LO STATUS EPISTEMOLOGICO DELLA CODIFICA TESTUALE.....	66
LA CODIFICA COME LINGUAGGIO TEORICO .....	73
<b>LA RAPPRESENTAZIONE COMPLESSA DI TESTI COMPLESSI</b> ..	<b>77</b>
IL PROBLEMA DELLE GERARCHIE SOVRAPPOSTE .....	80
UNA GERARCHIA PER LE GERARCHIE SOVRAPPOSTE .....	82
SOLUZIONI PER IL PROBLEMA DELLE GERARCHIE SOVRAPPOSTE.....	84
LE SOLUZIONI INTERNE AL PARADIGMA XML .....	85
OLTRE XML COSÌ COME LO CONOSCIAMO .....	91
CONCLUSIONI .....	96
<b>MODELLI E METODI DI ANALISI COMPUTAZIONALE DEL</b>	
<b>TESTO</b> .....	<b>97</b>

**TEMPI VERBALI E STRUTTURA NARRATIVA: UN MODELLO DI  
ANLISI COMPUTAZIONALE DEL TESTO NARRATIVO..... 103**

L'ATTEGGIAMENTO LINGUISTICO E IL PUNTO DI VISTA .....	108
LA PROSPETTIVA LINGUISTICA E GLI UNIVERSI TEMPORALI NELLA NARRAZIONE .....	111
IL RILIEVO NARRATIVO NEL TESTO .....	114
LE TRANSIZIONI TEMPORALI E LA COERENZA TESTUALE .....	116
CONCLUSIONI .....	118

***IL TURNO DI PIRANDELLO: UN ESPERIMENTO DI ANALISI  
ASSISTITA DAL COMPUTER..... 121***

TEMPI VERBALI E CONFIGURAZIONE ENUNCIATIVA DEL ROMANZO .....	123
LA PROSPETTIVA LINGUISTICA: TEMPORALITÀ PROGRESSIVA DELL'INTRECCIO E IRRAZIONALITÀ DEGLI EVENTI .....	128
RILIEVO NARRATIVO E TIPOLOGIA DEI PROCEDIMENTI NARRATIVI .....	130
DISCORSO INDIRETTO LIBERO E SOLILOQUIO DEI PERSONAGGI .....	134
DISPOSIZIONE SINTAGMATICA DEI TEMPI E SEGMENTAZIONE DEL DISCORSO NARRATIVO .....	136
IL TURNO E LA CRISI DEL NATURALISMO .....	140

**BIBLIOGRAFIA ..... 147**

## Introduzione

Lo studio della letteratura, della filosofia, della storia, e in generale gli studi umanistici hanno come oggetto privilegiato i testi. Intorno a questo oggetto di studio, ciascuna disciplina ha sviluppato nel corso dei secoli un insieme di procedimenti di trattamento e metodologie di analisi. Molti studiosi, seguendo e sviluppando le analisi e le intuizioni di Marshall McLuhan e di Walter Ong, hanno mostrato come queste pratiche sono state fortemente determinate dalle tecnologie di produzione e diffusione dei supporti fisici dei testi che si sono storicamente succeduti: rotolo di papiro, codice manoscritto, libro a stampa.

La diffusione delle tecnologie informatiche e telematiche a cui abbiamo assistito negli ultimi anni sta determinando una nuova rivoluzione nelle tecnologie di produzione e trasmissione del sapere. Come già avvenuto nel passato, il passaggio dall'era di Gutenberg a quella di Turing e von Neuman determina inevitabilmente profonde trasformazioni nelle pratiche sociali e scientifiche che hanno per oggetto o fine i testi<sup>1</sup>.

Il testo, trasferito su un supporto costituito da astratte unità di informazione, diventa testo elettronico o digitale. Una importante conseguenza di questa trasformazione è l'incontro tra ambiti disciplinari che, nel senso comune, non diremo dell'uomo della strada, ma anche dei rispettivi addetti ai lavori, apparivano, fino a pochissimo tempo addietro, irriducibilmente incompatibili: il mondo della ricerca letteraria e quello delle scienze e tecnologie dell'informazione. Un incontro che, ben lungi dal limitarsi a un puro e semplice rapporto strumentale, produce una vera e propria interazione tra saperi diversi, ognuno dei

---

<sup>1</sup> Turing è il matematico inglese che ideò, per fini logico-matematici, una classe di macchine ideali, detta appunto "macchine di Turing", che è il modello astratto alla base di qualsiasi macchina computazionale, quali sono i *computer* digitali. Il modello vero e proprio di macchina digitale a cui questi si rifanno fu invece elaborato da John von Neuman, altro geniale matematico di origine ungherese ma vissuto negli Stati Uniti, a cui si deve in gran parte il progetto del primo computer digitale.

quali fornisce prospettive differenti agli altri, sia sul piano degli strumenti e delle metodologie, sia su quello dei contenuti.

Un settore nel quale la produttività di questa interazione si è resa particolarmente evidente, non casualmente, è lo studio dei problemi connessi alla codifica informatica dei testi, che è stata ed è tuttora al centro di una notevole attenzione da parte della comunità scientifica. Sarebbe sufficiente una semplice rassegna bibliografica dei lavori pubblicati nel campo a partire dagli anni '80 per suffragare questa osservazione con dati oggettivi, e per fugare l'impressione che il punto di vista individuale di chi scrive, in ragione della lunga frequentazione con l'argomento (si è persino parlato di una scuola romana della codifica), possa in un certo senso sovrastimare la rilevanza del tema. Ma anche limitandosi a scorrere l'indice della più importante autorevole e completa (ancorché criticabile da molti punti di vista) sinossi enciclopedica dedicata all'Informatica umanistica, il *Companion to Digital Humanities* della Blackwell (recentissimamente pubblicato in forma gratuita sul Web, <http://www.digitalhumanities.org/companion/>), ci si può confermare in questa convinzione.

Cosa intendiamo, dunque, con “codifica informatica dei testi”? Intuitivamente, la codifica è la rappresentazione dei testi su un supporto digitale in un formato utilizzabile da un elaboratore. In linea generale questa rappresentazione solleva due ordini di problemi.

Il primo ordine di problemi è di natura squisitamente tecnica, e riguarda le tecnologie e i sistemi di riproduzione e memorizzazione elettronica dei testi. Se le questioni teoriche si offrono all'attenzione degli umanisti con maggiore fascino, i temi più strettamente tecnici non vanno assolutamente sottovalutati. Si tratta infatti di individuare le strategie ottimali per effettuare il passaggio dal predominio ormai secolare della stampa alla comunicazione del sapere con tecnologie digitali.

I nuovi media informatici sono dotati di grande capacità conservativa, versatilità nel rappresentare fenomeni comunicativi i più diversi, e capacità di diffusione. Ma, se da un lato il medium digitale amplifica l'accessibilità e la pervasività delle informazioni e del sapere, dall'altro esso induce a eliminare tutto ciò che non rientra nei limiti imposti dalla sua natura formale. Per limitare questo rischio è necessario far intervenire le competenze diversificate che si muovono intorno ai fenomeni della comunicazione, comprese quelle degli umanisti. In caso contrario quella parte del patrimonio intellettuale collettivo, del sapere, che è in una forma non facilmente disponibile alla traduzione

elettronica, rischia di scomparire dalla comunicazione sociale o di essere irrimediabilmente distorto.

Il secondo ordine di problemi concerne l'epistemologia e l'ontologia delle scienze del testo e la natura dei loro oggetti di studio. Il termine "testo" infatti si riferisce a un oggetto complesso e plurale, in cui esiste un livello materiale, il supporto e le tracce d'inchiostro, un livello astratto, la sequenza verbale, la quale a sua volta genera una serie di livelli di contenuti semantici. Ognuna di questi livelli testuali sono oggetto di altrettante discipline e/o teorie che si occupano dei testi. Esiste dunque un rapporto di determinazione reciproca tra la molteplice definizione della natura del testo e i linguaggi con i quali descriviamo e teorizziamo sul testo. Se consideriamo la rappresentazione informatica come una forma speciale di linguaggio teorico, la valenza di questa osservazione può essere estesa anche alla codifica.

Questo ci permette di riconsiderare la nostra intuitiva definizione, precisando che la codifica è la rappresentazione formale di un testo a un qualche livello descrittivo. Si tratta dunque di un processo rappresentazionale che implica una serie operazioni di selezione e classificazione degli elementi rilevanti in funzione di un determinato punto di vista. Tali operazioni si applicano a diversi livelli descrittivi: si può attribuire una data porzione "x" di un dato documento testuale alla classe dei "titoli di capitolo"; oppure si può attribuire un dato segmento di un testo narrativo alla funzione narratologica "complicazione". Siamo comunque in presenza di inferenze guidate da una teoria preesistente, o di abduzioni che ne istituiscono una: come ha efficacemente rilevato Sperberg-McQueen "la codifica riflette una teoria del testo".

Ciascun atto interpretativo è caratterizzato da diversi gradi di certezza. Ma anche le inferenze che sembrano più solide e condivise sono sottoposte a norme interpretative che possono essere revocate in dubbio da nuove teorie, generali o settoriali. Basti pensare alla difficoltà nel lavoro filologico di attribuire una certa traccia su un manoscritto a un certo carattere, o di sciogliere certe forme di abbreviazione.

Si noti che questa visione pluralista non comporta necessariamente l'accettazione di una deriva "relativista". Al contrario la codifica informatica del testo permette allo studioso del testo di avvalersi di un linguaggio formale per esporre una data interpretazione del testo in forma esplicita, controllabile e intersoggettiva. Alcuni studiosi dei problemi della codifica hanno usato la nozione di modello, mutuandola dall'epistemologia scientifica, per caratterizzare teoricamente la

rappresentazione testuale. Ogni codifica testuale produce un modello del testo. Alcuni di questi modelli possono essere messi in relazione tra loro, in modo che un processo computazionale sia in grado di trasformare un modello A nel modello B. La codifica informatica dei testi dunque permette alle discipline letterarie di dotarsi di un sistema di verifica sperimentale per le loro interpretazioni e teorie sui testi.

Possiamo ora porci una domanda: la critica letteraria viene modificata e arricchita dall'utilizzazione dell'informatica, le due macchine argomentative si incontrano e si modificano a vicenda, oppure l'attività critica rimane sostanzialmente la stessa pur avvalendosi "d'una penna che scrive più veloce"? La domanda, posta in altri termini ancora, può diventare: esiste uno specifico della critica letteraria informatica che giustifichi, al di là delle potenzialità tecnologiche offerte, l'impegno non solo nell'apprendimento delle tecniche, ma anche nella riflessione teorica in questo settore.

Crediamo che tale specifico sia ancora una volta tutto di natura metodologica e consista nella sfida alla riduzione dell'implicito. È questa la sfida che il critico e il teorico accettano nel momento in cui decidono di adottare gli automi computazionali per studiare un testo. L'automatizzazione di procedure analitiche riduce lo spazio che l'analisi tradizionale lasciava all'indefinito. Il ricercatore, in un certo senso, rinuncia a quello che sa e accetta di riprogettare e ricostruire il proprio "sapere".

Ci si sposta, quindi, da uno spazio nel quale la procedura critica era scarsamente definita e documentata a una fase in cui il lavoro scientifico nel settore letterario deve produrre una esaustiva documentazione la quale, garantendo la conservazione della maggior parte dell'informazione contenuta nel testo di partenza (o la sua fruttuosa trasformazione) ed esplicitando dati e metodi, avvicini la ricerca letteraria a quella che più comunemente viene definita ricerca scientifica: esplicitati i dati, rese note le procedure, qualunque ricercatore deve poter arrivare ai medesimi risultati.

In questo orizzonte tematico e in questa prospettiva teorica si collocano i contributi qui raccolti e rielaborati, prodotti dall'autore nel corso di un decennio di ricerca e sperimentazione.

## Codifica e modellizzazione del testo

Lo studio della letteratura ha come oggetto immediato i testi letterari. Nella nostra cultura la quasi totalità dei testi letterari (fino a questo momento) è costituita da testi scritti su supporti cartacei di varia natura e forma. Poiché, allo stato dell'arte, i sistemi informatici non sono in grado di estrarre dati da tale tipo di supporto, l'introduzione dell'informatica nello studio letterario comporta l'esigenza di rendere disponibili i testi della letteratura su supporti elettronici.

La memorizzazione informatica del testo risponde appunto a questa esigenza. Possiamo dunque definire, in via preliminare e generale, come memorizzazione di un testo la trasposizione completa esaustiva e fedele del testo su un supporto materiale (disco magnetico, nastro, disco ottico, schede di memoria) e in una forma tale da renderne possibile il caricamento nella memoria del computer e il trattamento informatico.

Questa definizione preliminare, tuttavia, costituisce solo una delimitazione pragmatica del processo che abbiamo denominato memorizzazione, e non ha alcuna pretesa teorica: si tratta di una base di partenza per approfondire la riflessione.

D'altronde, è oramai opinione condivisa nell'ambito dell'informatica umanistica che la memorizzazione del testo non è riducibile a una semplice e meccanica trasposizione. A una più attenta e profonda analisi, essa si rivela un complesso processo comunicativo, una trasmissione di informazioni, a cui sono sottesi fenomeni semiotici. Infatti, astruendo dal circuito comunicativo complessivo, in cui destinatario è il fruitore del testo (lettore o critico) e destinatario il suo autore, ci troviamo in presenza di una fonte di informazioni, il testo, che un emittente/codificatore deve inviare verso un destinatario, la memoria del calcolatore, attraverso un canale, l'unità di input utilizzata, veicolate da segnali elettronici. Come tutti i processi comunicativi, la trasmissione delle informazioni testuali deve essere mediata da un dispositivo che renda possibile la comunicazione tra i due attori del processo, ovvero da un codice.

Il processo di memorizzazione elettronica può in questo modo, essere ricondotto al problema più generale di costruzione di uno o più codici di rappresentazione dell'informazione testuale, che descrivano tutti gli elementi significanti individuati dal codificatore nella fonte del processo comunicativo (il testo), e che si adattino alla natura fisica del canale e del destinatario (il computer). Con il concetto di codice intendiamo qui, seguendo la definizione che ne è stata data in ambito semiotico, un sistema convenzionale di correlazioni tra due sottosistemi: uno costituisce il sistema delle unità *in praesentia*, i segnali, in un atto comunicativo, (quello che Hjelmslev ha chiamato piano dell'espressione); l'altro il sistema delle unità *in absentia*, ovvero il piano del contenuto, o delle unità veicolate. Dunque costruire un codice per memorizzare i testi significa:

- individuare correttamente ed esaustivamente i dati pertinenti della fonte (sistema semantico del codice);
- progettare un sistema di correlazioni biunivoche tra tali elementi e unità espressive discrete, compatibili con la natura fisica del ricevente elettronico.

Evidentemente il nodo problematico è rappresentato dal conseguimento di una codifica adeguata delle informazioni della fonte. Si dice problematico perché la fonte (il testo) è a sua volta una unità informazionale complessa, alla cui costituzione cooperano molteplici codici, non rigidamente calcolabili (a partire dalla codifica alfabetica, per arrivare agli elementi visivi e materiali che costituiscono un testo a stampa), mentre le caratteristiche fisiche di un processore automatico di informazioni impongono che il codice di comunicazione verso la macchina sia, in ultima analisi, il codice di numerazione binaria. Esso è composto di soli due segni, "1" vs "0" e da una sintassi che impone la logica ferrea del terzo escluso, per cui un evento o si dà, senza ambiguità, o non si dà, *tertium non datur*: non è possibile prevedere situazioni intermedie, accadimenti parzialmente riconoscibili o variamente interpretabili. A questa struttura rigida del codice nell'ambito informatico va confrontata la complessa intersezione di codici che costituiscono un testo da memorizzare.

## Codici e modelli

Un aspetto di notevole interesse nelle proprietà generali dei codici, legato con la sua funzione comunicativa, che è stato messo in rilievo dalla semiotica slava, in particolare da Lotman [1970], è la funzione di modellizzazione della realtà svolta dal sistema semantico di un codice.

Infatti, come rileva il grande studioso recentemente scomparso, supponiamo di avere a disposizione, per descrivere (parlare di) una certa porzione del mondo, una lingua composta di due soli segni: è evidente che questa lingua riflette un certo modello di quella porzione di mondo, riducendone la complessità, attraverso un processo di astrazione e pertinentizzazione, in due classi semantiche. Inversamente si potrebbe dire che, in presenza di una realtà complessa (che contingentemente può risultare strutturata in livelli semiotici, essendo in parte prodotto di attività semiotica), costruire un modello di tale realtà significa darne una descrizione codificata in un linguaggio.

Non possiamo in questa sede approfondire le implicazioni teoriche e filosofiche di queste osservazioni. Tuttavia esse ci consentono di guardare al problema della codifica elettronica dei testi da un punto di vista molto interessante, utilizzando la potenzialità euristica del concetto di modello, e di alcune proprietà di un modello che sono state studiate in sede epistemologica e cognitivista. Pur riconoscendo la necessità di procedere con estrema cautela allorché si trasportino strumenti teorici in ambiti e domini disciplinari diversi da quelli in cui sono stati teorizzati, la definizione di modello scientifico ha innegabili parallelismi con la nozione di modello che abbiamo tracciato sopra. Ad esempio Rosenblueth e Wiener, due tra i fondatori della cibernetica, affermano:

Nessuna parte costitutiva dell'universo è talmente semplice da lasciarsi afferare e controllare senza astrazione. Quest'ultima consiste nel prendere la porzione di universo considerata e nel sostituirla con un modello avente struttura analoga, ma più semplice. [Rosenblueth, Wiener, 1945 (1986: 76)]

In ambito epistemologico e scientifico i modelli sono suddivisi in due classi: modelli materiali, ovvero riproduzioni fisiche del dominio modellizzato che rispondono a criteri di similitudine strutturale e funzionale, e modelli formali o astratti. In questa sede ci riferiremo in particolare al concetto di modello formale. Un modello formale è una descrizione simbolica di un oggetto o di un fenomeno, che riproduce

le relazioni osservate nella realtà modellizzata con un livello più o meno alto di astrazione, e che deve fornire le regole per effettuare predizioni su tale realtà, come rileva Marvin Minsky:

Se una creatura è in grado di rispondere a una domanda relativa a un esperimento ipotetico senza realizzarlo effettivamente, si dice che essa dimostra di avere una certa conoscenza del mondo. La sua risposta alla domanda in questione non è altro, infatti, che una descrizione codificata del comportamento (avente luogo al suo interno) di qualche sottomacchina o “modello” rispondente a una descrizione codificata della situazione del mondo descritta dalla domanda stessa. [Minsky, 1968 (1986: 307)]

Si può dire che un modello formale è una rappresentazione fatta attraverso un linguaggio formale (matematico, logico-formale, insiemistico, etc.), i cui elementi hanno un ruolo funzionale omologo a quello dei corrispondenti elementi individuati nella realtà modellizzata. Anche gli algoritmi, per inciso, sono modelli di realtà espressi in linguaggi formali; più precisamente si tratta di modelli prescrittivi: essi indicano la serie finita di operazioni da compiere per ottenere un certo risultato, simulando una certa realtà fenomenica.

I modelli, sia formali sia materiali, hanno due proprietà molto importanti. La prima, che ha un fondamento gnoseologico, è l'impossibilità teorica di conseguire la coincidenza biunivoca tra il modello e la realtà che descrive, come ci insegna il ben noto paradosso della mappa di un paese descritto da Lewis Carrol in Sylvie e Bruno: un tale modello, o coinciderebbe con la stessa realtà che modella, o dovrebbe essere formulato da un ente che coglie la realtà nella sua interezza immediatamente (il grado di conoscenza “intuitiva e adeguata” che Leibniz, attribuiva al divino).

La seconda proprietà di un modello, che ha delle conseguenze fondamentali, è la sua variabilità rispetto al punto di vista dell'osservatore, che viene così enunciata dallo stesso Minsky:

... per un osservatore B un oggetto A\* costituisce un modello di un oggetto A nella misura in cui può usare A\* per rispondere alle domande sul conto di A che per lui rivestono interesse. La relazione che definisce il modello è intrinsecamente ternaria. Ogni tentativo di abolire il ruolo svolto dalle intenzioni del ricercatore B conduce a definizioni circolari o ad ambiguità circa “caratteristiche essenziali” e simili. [Minsky, 1968 (1986: 307)]

Ogni modellizzazione della realtà è, dunque, un processo triadico, che comporta la interrelazione tra osservatore, oggetto e modello: date

le condizioni iniziali, si può dire che esiste una precisa relazione funzionale che lega punto di vista dell'osservatore, oggetto e modello: questo ultimo risulta funzione del punto di vista dell'osservatore sull'oggetto.

La prima conseguenza di questa caratteristica fondamentale è la *intenzionalità* della modellizzazione, ovvero il suo dipendere dalla volontà dell'osservatore di scegliere un determinato punto di vista o un altro. Una seconda conseguenza, anche questa di notevole rilievo ai fini del discorso che stiamo conducendo, è la possibilità teorica che possano esistere infiniti modelli della realtà, corrispondenti alla infinita variabilità dei punti vista che un osservatore può assumere.

Naturalmente si possono definire dei criteri pragmatici per valutare la validità o meno di una modellizzazione. In via provvisoria possiamo dire la determinazione dell'adeguatezza di un modello, e dunque di una codifica, va misurata su questi due parametri:

- esaustività in rapporto al punto di vista che si è assunto sull'oggetto;
- formalizzabilità, ovvero possibilità di tradurlo in un linguaggio formale computabile.

Il primo punto rimanda al problema della definizione chiara e strutturata degli elementi che costituiscono un oggetto, a un determinato livello, e delle relazioni tra questi elementi. Il secondo punto invece comporta la scelta o la progettazione di un linguaggio descrittivo dell'oggetto in relazione al punto di vista scelto, che sia o il codice binario, o una trasformazione definita del codice binario, poiché il computer è in grado di comprendere solo tale famiglia di linguaggi.

### **Testo e modelli del testo**

Alla luce di queste riflessioni sul concetto di modello, si può osservare come la codifica elettronica di un testo possa essere descritta come un processo di modellizzazione. Infatti per poter attuare la codifica del testo occorre costruire uno o più modelli descrittivi di tale testo, implementati in un linguaggio comprensibile alla macchina (dove comprensibile sta per processabile). Dalla precedente discussione risulta evidente che ogni modellizzazione implica necessariamente la scelta di un punto di vista da parte dell'osservatore/codificatore, non

essendo nemmeno teoricamente pensabile un modello unico e assoluto che descriva esaustivamente un testo.

In conseguenza di questa impossibilità teorica, la scelta di un modello per la codifica elettronica da parte dello studioso, non potendo mai essere assolutamente neutrale, sarà, pragmaticamente, legata alle esigenze poste dalla ricerca. Diversi modelli potranno, infatti, descrivere un testo, con diversi gradi di interpretazione rispetto a strutture presenti, o ipotizzabili, nell'oggetto, ognuno dei quali in grado di soddisfare particolari problemi di ricerca, purché rispondano ai criteri di adeguatezza che abbiamo enunciato sopra.

Finora abbiamo utilizzato il termine "testo" in modo acritico, per designare l'oggetto della memorizzazione, affermando che si tratta di una entità informativa complessa. Negli ultimi anni sono state elaborate molteplici teorie del testo in diversi ambiti disciplinari, tutte fondate su modelli multiplanari. Ma, nella maggior parte dei casi esse forniscono una definizione del concetto di testo basata esclusivamente sul livello verbale (si veda la grande massa di lavori teorici sviluppati nell'ambito della *Textlinguistik*), spesso sul versante esclusivamente semantico di tale aspetto (coerenza, pronominalizzazione, thema-rema, ecc.).

Tuttavia, quando si traspone un testo informato digitale, l'operazione che si attua è in effetti la memorizzazione del testo per come esso appare in un particolare libro, singolo esemplare di una determinata edizione di un determinato o indeterminato manoscritto che supportava quella che l'autore, in un dato momento del suo periodo creativo, considerava una redazione definitiva dell'opera. Ognuno degli infiniti atti di fruizione di un testo (e dunque di riproduzione elettronica), non può prescindere da una mediazione materiale che rende presente il testo stesso. A questo proposito ci sembra assai feconda una breve osservazione di Cesare Segre:

Se consideriamo i segni grafici (lettere, interpunzione, ecc.) come significati di suoni, pause, ecc., e riflettiamo sul fatto che questi segni possono essere trascritti più volte e in vari modi (per esempio con grafia e caratteri diversi), restandone immutato il valore, possiamo concludere che il testo è l'invariante, la successione di valori, rispetto alle variabili dei caratteri, della scrittura, ecc. Possiamo anche parlare di significati, se si precisa che si allude a significati grafici, quelli della serie di lettere e segni d'interpunzione che costituiscono il testo. Il testo è dunque una successione fissa di significati grafici. [Segre, 1985: 29]

Da questo punto di vista possiamo definire il testo come una entità astratta invariante, che in ogni operazione di realizzazione materiale della sequenza di simboli grafici determina la struttura fisica di un oggetto sensibilmente concreto (ovvero capace di attivare uno dei canali recettivi dell'uomo verso stimoli esterni), che costituisce il supporto materiale, stabile e riproducibile dell'informazione testuale. Distingueremo questo oggetto dal testo chiamandolo documento, in riferimento alla definizione datane da Robert Escarpit:

Si definirà dunque il documento come un oggetto informazionale visibile o toccabile, dotato di una doppia indipendenza rispetto al tempo:  
- sincronia: indipendenza interna del messaggio, che non è più una sequenza lineare di eventi, ma una giustapposizione multidimensionale di tracce;  
- stabilità: indipendenza globale dell'oggetto informazionale, che non è più un evento iscritto nello scorrere del tempo, ma un supporto materiale della traccia, che può essere conservato, trasportato, riprodotto. [Escarpit, 1976 (1979: 144)]

Un documento testuale, a sua volta, non è solo un oggetto fisico, e nella sua costituzione intervengono altri fattori di carattere semiotico che contribuiscono in maniera più o meno diretta alla fruizione del testo. Infatti, come rileva Umberto Eco (riferendosi a un manoscritto, ma il discorso è applicabile anche al testo a stampa) ogni livello materiale può costituire espressione o contenuto di un sistema semiotico a seconda dei punti di vista da cui viene osservato:

... la scrittura rappresenta la sostanza linguistica, l'inchiostro è il supporto della manifestazione grafemica (vista come una forma), la pergamena è il supporto della disposizione dell'inchiostro (vista come una forma), i tratti fisico-chimici della pergamena sono il supporto delle sue qualità formali, e così di seguito. [Eco, 1990: 184]

L'opposizione testo vs documento individua i due livelli, o meglio punti di vista sul nostro oggetto di studio, pertinenti per effettuare una memorizzazione elettronica del testo con fini editoriali. Distingueremo, dunque, un modello elettronico del testo e un modello elettronico del documento originale.

Quando si dice modello del testo, occorre ricordare, non ci si riferisce qui ai fenomeni sintattici o semantici verbali o ultra-verbali, ma al testo come invariante di significati grafici. Tuttavia oltre alla sequenza di grafemi, a un testo nel senso da noi indicato possono essere ascritte anche le segmentazioni logiche e le partizioni interne di interi blocchi

di scrittura, che hanno una certa distribuzione gerarchica e funzionale, e che hanno una influenza diretta sulla composizione tipografica del documento: ad esempio un capitolo viene identificato da uno spazio bianco, da un salto pagina, da un titolo, o un identificatore numerico, etc. Questi elementi, infatti, possono essere descritti indipendentemente dall'interpretazione del significato semantico del livello verbale, malgrado questo non sia sempre vero: esistono moltissimi casi, specialmente in edizioni non contemporanee, in cui la partizione in capitoli non esiste nel manoscritto o negli incunaboli più antichi, essendo stata introdotta solo più tardi, per interpretazione dello stampatore. Ciò è dovuto naturalmente alla complessità dei rapporti semiotici tra livelli: il testo ha una certa struttura, i cui elementi sono determinati dalla struttura logico semantica del discorso, a sua volta legata al livello dell'intreccio, se si tratta di un romanzo, o comunque della sequenza dei motivi.

Sono ascrivibili alla struttura astratta di un testo anche alcuni elementi che Gerard Genette [1989] ha definito in un suo recente ponderoso volume dal suggestivo titolo di *Soglie*, peritesto editoriale: il titolo e il nome dell'autore, astraendo dalla loro collocazione spaziale, che appartiene al documento, costituiscono gli identificatori del testo, così come gli intertitoli, o le note. In generale possiamo dire che tutto ciò che è riconducibile alla istanza di responsabilità dell'autore possa far parte della struttura astratta del testo, anche se questa assunzione necessiterà di ulteriori approfondimenti, poiché possono darsi molteplici eccezioni che un modello deve essere in grado di recepire. Ogni modello come detto è risultato di una astrazione, e dunque necessariamente non può tenere conto dell'intera rete di rapporti e relazioni esistente in un oggetto del mondo, ma deve essere in grado di modificarsi, deve cioè, per usare un termine della epistemologia, essere un modello flessibile.

Passiamo ora al modello del documento. Esso deve riprodurre la struttura spaziale del supporto originale del testo, e la disposizione e costituzione delle tracce fisiche dei grafemi. Un tale modello ha una importante funzione scientifica, in primo luogo perché esso potrebbe essere oggetto di studio in sé per discipline quali la bibliografia, la codicologia, l'ecdotica, la storia delle tecnologie comunicazionali. In secondo luogo, una tradizione secolare, che nasce nel periodo alessandrino e giunge attraverso i *carmina intexta* medievali, i *carmina figurata* barocchi, alle avanguardie letterarie del nostro secolo, attesta come molti autori hanno avuto presenti le possibilità significative che

derivano da una semantizzazione esplicita degli elementi non verbali di un testo scritto.

Si pensi alla rilevanza che ha questo ambito di problemi nella redazione dei testi poetici, e alla tematizzazione esplicita che di ciò le avanguardie del nostro secolo hanno fatto, a partire dal *Coup de dés* di Mallarmé, fino alle più recenti esperienze di poesia visiva e poesia concreta. Ma anche nella storia del romanzo vi sono esempi di consapevolezza del ruolo della disposizione tipografica e topografica del segno grafico nella pagina bianca, come il *Pantagruel* di Rabelais, il *Tristram Shandy* di Sterne, *Alice nel paese delle meraviglie* di Lewis Carroll.

### Modelli del testo e codifica digitale

Possiamo riassumere in maniera schematica gli elementi pertinenti che devono costituire il modello del testo e il modello del documento nel seguente quadro sinottico:

Testo	Documento
sequenza di grafemi	materializzazione di una sequenza astratta di significati grafemici
sequenza strutturata di blocchi di testo con funzioni logiche differenziate	composizione tipografica delle linee di scrittura, delle immagini, dei caratteri
elementi paratestuali discorsivi	organizzazione e struttura del supporto materiale

Si tratta ora di individuare dei linguaggi computabili per costruire effettivamente tali modelli elettronici. La diffusione dei personal computer e le esigenze di elaborazione e formattazione dei testi in formato elettronico ha portato allo sviluppo di molteplici sistemi orientati alla creazione e manipolazione di testi. Tra questi hanno una notevole importanza ai fini della modellizzazione di testi, quei sistemi basati sui cosiddetti *markup language*.

Il termine inglese *markup* designava nella stampa tipografica tutte le indicazioni e annotazioni simboliche aggiunte dall'autore o dall'editore su un manoscritto o su un dattiloscritto per istruire il tipografo o il compositore relativamente all'impaginazione, alla scelta dei caratteri, delle spaziature e così via (ad esempio la sottolineatura per indicare che il testo deve essere stampato con il carattere corsivo). Si-

milmente un *markup language* è costituito da un set di istruzioni di un vero e proprio linguaggio orientato alla descrizioni dei fenomeni di composizione e struttura del testo.

Altrettanto, se non maggiormente complessa e problematica, è la modellizzazione del documento. Infatti rappresentare la struttura fisica di un documento, che viene percepita da un osservatore come una struttura analogica, attraverso dei linguaggi informatici di alto livello, con una esaustività adeguata da permettere studi codicologici o bibliografici, risulta al momento impossibile. D'altra parte un tale modello dovrebbe sostanzialmente effettuare una traduzione linguistica di una serie di codici visuali e iconografici, e questa operazione comporterà sempre una perdita di informazione. Crediamo che una adeguata soluzione pragmatica di questi problemi sia da individuare nell'associazione di un modello implementato da un markup language, fin dove è possibile, e di un modello grafico del documento, implementato in uno dei formati grafici standard, tra quelli sviluppati nell'ambito delle tecnologie di computer grafica.

Questi linguaggi grafici, grazie allo sviluppo delle tecnologie hardware, sono in grado di fornire rappresentazioni di immagini con un elevatissimo grado di definizione, anche senza avere a disposizione stazioni hardware di potenza e costi proibitivi per un singolo utente di personal computer, e questi standard di qualità tenderanno inevitabilmente a crescere inversamente ai costi con il procedere dello sviluppo tecnologico. Una tale forma di modellizzazione informatica di un documento testuale risulterebbe, peraltro, la più adeguata nel caso specifico dei manoscritti, per i quali nessuna descrizione di tipo linguistico sarebbe in grado di rappresentare tutte le informazioni visuali che una immagine digitale è in grado di veicolare.

## I linguaggi per la codifica del testo

L'applicazione di tecnologie informatiche nelle discipline umanistiche, e in particolare nelle scienze letterarie, ha come suo fondamento la rappresentazione degli oggetti che costituiscono il loro dominio di studio. Tali oggetti sono in linea generale i testi<sup>1</sup>. Nell'ambito di queste discipline i testi vengono sottoposti a una variegata tipologia di trattamenti: scrittura (il momento in cui il testo ha origine), edizione, lettura, analisi, interpretazione, archiviazione, catalogazione. Dunque l'applicazione di procedure informatiche al trattamento dei testi richiede anche la simulazione dei processi che su di essi vengono effettuati.

In campo informatico il problema della rappresentazione di dati e processi è stato studiato con molta attenzione. La natura dei processi computazionali e dei linguaggi con i quali essi vengono specificati, infatti, determina una serie di restrizioni sulla forma di queste rappresentazioni. In primo luogo occorre ricordare che esse sono sostanzialmente rappresentazioni formalizzate. Un computer infatti è un tipo particolare di macchina formale automatica che manipola simboli secondo regole per effettuare procedimenti di calcolo. Il linguaggio dell'elaboratore è un linguaggio formale costituito da un alfabeto finito di simboli, e da regole di manipolazione di questi simboli: la macchina "capisce" solo formule ben formate espresse in questo linguaggio<sup>2</sup>. Si deve notare che «ogni dato su cui l'elaboratore deve operare viene rappresentato a livello elementare mediante una sequenza (o stringa) di simboli. Anche il programma, cioè la sequenza di istruzioni

---

<sup>1</sup> Rientrano tra gli oggetti delle scienze letterarie anche entità e fenomeni assai complessi come ad esempio i generi, i movimenti letterari, o l'intertestualità. Nondimeno questi sono fenomeni ed entità derivate dai testi. Ciò comporta che la loro rappresentazione informatica non può prescindere da una rappresentazione adeguata delle unità fondamentali del dominio stesso, appunto i testi.

<sup>2</sup> Naturalmente in questo contesto occorre dare il giusto significato al termine "capire": si tratta di un capire *esclusivamente sintattico*, nel senso logico del termine, e non di una interpretazione semantica.

che l'elaboratore deve interpretare ed eseguire, è codificato con una stringa di simboli» [Ausiello et al., 1991: 22-23].

La specificazione di un sistema di codifica è il primo passo nella costruzione di un sistema informativo automatizzato. In generale in informatica si distinguono i formalismi che servono a rappresentare dati da quelli che servono a rappresentare processi. I primi sono indicati come strutture o tipi di dati, i secondi come linguaggi di programmazione. Naturalmente questi due tipi di formalismi sono spesso strettamente legati, e i tipi di dati fanno parte delle primitive di ogni linguaggio di programmazione.

La struttura la complessità e l'espressività dei formalismi informatici può variare notevolmente, lungo una scala che ha come unità di misura il livello di astrazione rispetto ai processi fisici che avvengono nel calcolatore e che corrispondono alle sue computazioni fondamentali.

È ben noto che a un estremo di questa scala, il livello più vicino alla macchina per così dire, ci sono la notazione binaria e le operazioni booleane sui simboli binari, che permettono di effettuare le operazioni aritmetiche fondamentali. All'altro estremo troviamo invece una vasta congerie di linguaggi di programmazione o di specificazione di strutture dati che usano simbolismi formalizzati molto astratti rispetto ai processi computazionali e che in taluni casi cercano di modellizzare i processi di astrazione e di ragionamento umano.

Alla luce di queste considerazioni possiamo dire che l'applicazione di metodologie computazionali nell'ambito della ricerca umanistica comporta due tipi, o meglio due fasi di formalizzazione:

- definizione e implementazione di strutture dati adeguate alla cattura dei fenomeni di interesse dell'umanista, e in particolare alla rappresentazione formale dei testi;
- specificazione di algoritmi che, applicati alle strutture dati, siano in grado di simulare i processi di manipolazione dei testi tipici della ricerca umanistica o in generale delle pratiche sociali che hanno a che fare in vario modo con i testi.

Il problema della codifica testuale rientra in generale nel primo tipo di formalizzazione, anche se, come è stato evidenziato [Buzzetti, 1996], è assai difficile, e spesso inopportuno dividere nettamente la determinazione delle strutture dati da quella degli algoritmi.

## **Schemi e linguaggi di codifica**

Se, come già detto, la codifica richiede allo studioso un notevole sforzo di esplicitazione e formalizzazione del sapere implicito contenuto nel suo operare (ma anche nel semplice atto di lettura “ingenua”), non meno importante è la definizione e implementazione di un linguaggio formale che deve essere a un tempo processabile da un elaboratore e sufficientemente espressivo per rappresentare la complessità dell’oggetto testo. Definiamo una simile struttura dati uno schema di codifica.

Più precisamente uno schema di codifica associa un insieme di caratteristiche o elementi costituenti di un oggetto testuale a un insieme di simboli, e le relazioni tra gli elementi testuali a relazioni sintattiche tra i simboli. Esso è dotato di una sintassi che è data dalla sintassi del formalismo informatico su cui è basato, e una semantica che è data dall’insieme di caratteristiche o elementi testuali che permette di rappresentare. Se il formalismo su cui si fonda lo schema di codifica è un vero e proprio linguaggio informatico, specializzato per la rappresentazioni di dati testuali, si parla anche di linguaggio di codifica, sebbene questa formula sia più rigorosamente da riferirsi al formalismo in quanto tale. Infatti un linguaggio informatico (sia esso un linguaggio di programmazione o un linguaggio di codifica), non ha propriamente una semantica nel senso che abbiamo individuato sopra, o la ha solo per alcune sue primitive.

I linguaggi per la codifica testuale vengono denominati nella letteratura anglosassone *markup language*, linguaggi di marcatura. L’espressione deriva per analogia con la simbologia introdotta in un manoscritto o in una bozza di stampa per indicare le correzioni e i trattamenti editoriali, chiamata in inglese *mark up*. Anche i linguaggi di markup infatti, consistono di un insieme di simboli che vengono inseriti all’interno o accanto al testo verbale. Vedremo come l’analogia con le pratiche editoriali sia stata, in una prima fase, assai stretta, per poi dare luogo a un concetto di markup molto più astratto e potente.

## **I requisiti di uno schema di codifica per la ricerca umanistica**

L’informatica allo stato attuale è in grado di fornire allo studioso di scienze umane una serie di tecnologie per implementare dei sistemi di rappresentazione testuale. Prima di analizzare le caratteristiche di que-

sti sistemi è bene individuare alcuni requisiti generali che uno schema di codifica testuale dovrebbe auspicabilmente possedere per essere utilizzato in applicazioni scientifiche.

Si noti che questi requisiti sono auspicabili non solo nell'ambito della ricerca scientifica. Infatti, la diffusione sempre maggiore delle tecnologie informatiche nella comunicazione sociale rende oggi possibile lo sviluppo di una forma di editoria elettronica che assuma i nuovi media come supporti dell'attività di pubblicazione. La carenza di consapevolezza su questi argomenti, finora riscontrabile in gran parte delle esperienze di editoria elettronica, si concretizza in prodotti di scarso valore editoriale, che non si discostano troppo dal mercato dello entertainment elettronico, anche quando aspirano a un livello culturale alto, o che rischiano di essere completamente inutilizzabili nello spazio di pochissimi anni.

### *Potenza rappresentazionale*

Codificare un testo significa esplicitare i processi inferenziali effettuati da un interprete nella comprensione del testo stesso. Il principale requisito di uno schema di codifica, pertanto, è la capacità rappresentazionale che esso offre allo studioso. Uno schema di codifica dunque deve essere dotato di formalismi in grado di:

- rappresentare adeguatamente i differenti fenomeni testuali che vengono studiati da varie discipline;
- permettere di esprimere tali fenomeni secondo le diverse metodologie e impostazioni teoriche che sono state sviluppate nel contesto di ciascuna disciplina, senza indicare alcuna opzione per una di esse in particolare.

La potenza di uno schema di codifica è determinata dalla natura della notazione o linguaggio informatico utilizzato, e dalla ricchezza e plasticità della sua sintassi.

### *Documentazione del testo*

La documentazione del testo elettronico è un aspetto che spesso viene posto in secondo ordine nelle discussioni dedicate al tema della codifica testuale. Essa invece è un requisito fondamentale di uno schema di codifica usato in applicazioni di tipo scientifico: sia al fine

di permettere una adeguata catalogazione del testo elettronico all'interno di archivi o corpora disponibili su stazioni di lavoro locali oppure in ambienti di rete; sia per facilitare una corretta interpretazione della codifica da parte di utenti distanti nello spazio e nel tempo dal responsabile primario della stessa.

Una corretta documentazione del testo elettronico dovrebbe prevedere almeno le seguenti informazioni:

- individuazione del testo elettronico attraverso le sue determinazioni bibliografiche: titolo, autore, luogo e data di edizione, etc.
- certificazione della responsabilità editoriale del testo, anche nel caso in cui la codifica ha avuto diversi responsabili
- indicazione della fonte su supporto originale del testo per i testi elettronici che sono edizioni di testi già editi su stampa o trascrizioni di manoscritti
- documentazione accurata delle metodologie di rappresentazione dei vari fenomeni testuali, delle scelte di approcci disciplinari o teorici che permettano di interpretare correttamente i simboli usati nella codifica del testo, delle eventuali correzioni e modifiche introdotte rispetto alla fonte

Per i testi che vengono diffusi attraverso canali telematici aperti (quale la rete Internet) è di capitale importanza anche l'individuazione di un sistema di certificazione della originalità e autenticità del testo, nonché un supporto ai procedimenti di reperimento in un ambiente di rete. Al momento questo è uno dei problemi più dibattuti negli organismi tecnici che definiscono le tecnologie della rete Internet, ma la discussione di questi aspetti tecnici esula dal presente scritto<sup>3</sup>.

#### *Portabilità e riutilizzabilità*

Il testo è un dispositivo comunicativo, che ha un valore in quanto può essere fruito dal maggior numero di individui, senza limitazioni di

---

<sup>3</sup> Rinviamo alla documentazione presenta in rete presso il server del *W3 Consortium*, nelle pagine dedicate al sistema detto *Uniform Resource Number*: L'indirizzo di rete è <http://www.w3.org>

spazio e di tempo. I supporti testuali, dunque, debbono garantire e facilitare questa universalità dell'accesso al loro contenuto. Indichiamo questo requisito con il termine tecnico portabilità.

Paradossalmente un testo su supporto digitale, che apparentemente sembra possedere la massima riproducibilità tecnica di cui parlava Benjamin, rischia di divenire inutilizzabile nel giro di pochissimi anni a causa del rapido sviluppo delle tecnologie hardware e software. Se alcuni libri a stampa hanno tranquillamente doppiato i cinquecento anni, e alcuni manoscritti i duemila anni, un attuale disco ottico fisicamente non supererà i cento anni, mentre logicamente sarà obsoleto in cinque, e potenzialmente inutilizzabile in dieci. L'archiviazione su supporto digitale del patrimonio letterario e culturale delle culture mondiali deve misurarsi con questi problemi, e adottare degli schemi di codifica capaci di garantire la massima portabilità. Il requisito della portabilità di un testo elettronico, sebbene sia di carattere tecnico, ha dunque un rilievo fondante.

Una risorsa informativa digitale è portabile se è intercambiabile tra sistemi diversi, riutilizzabile in molteplici processi computazionali anche a distanza di tempo, e integrabile da ulteriori risorse informative omogenee [Burnard, 1995a: par. 1].

In generale, per conseguire questi obiettivi il sistema di codifica dell'informazione (dei testi nel nostro contesto specifico) deve godere delle seguenti proprietà:

- indipendenza dall'hardware, ovvero da una particolare architettura elaborativa (processore), da un particolare supporto digitale (disco magnetico, disco ottico, etc.), o da un particolare dispositivo o sistema di output (video, stampa);
- indipendenza dal software, sia sistemi operativi, sia applicazioni deputate alla creazione, analisi, manipolazione e visualizzazione di testi elettronici; i software commerciali usano, nella maggior parte dei casi formati di dati proprietari mutuamente incompatibili; inoltre come scrive Sperberg-McQueen essi non sono "mai assolutamente adeguati a tutti gli utenti; mai egualmente graditi a tutti gli utenti; troppo rapidamente obsoleti" [1994];
- indipendenza logica dalle applicazioni ovvero indipendenza semantica dello schema di codifica da un particolare processo applicativo, o da una particolare applicazione, quale ad esem-

pio la stampa o l'edizione o l'analisi morfosintattica; uno schema di codifica infatti dovrebbe consentire di utilizzare il testo in ogni tipo di applicazione computazionale attuale o potenziale<sup>4</sup>.

### *Standardizzazione*

Abbiamo visto come la portabilità richieda una serie di requisiti tecnici. Tuttavia affinché uno schema o linguaggio di codifica sia sostanzialmente portabile è necessario che esso sia adottato universalmente dalla comunità di studiosi interessata teoricamente al suo utilizzo. Esso deve divenire uno standard.

In un senso stretto e formale uno standard è un insieme di norme relative a una particolare tecnologia che vengono emesse da un ente istituzionale nazionale o internazionale. L'ente internazionale deputato alla definizione di standard formali è la *International Standardization Organization*, composta da membri dei vari enti di standardizzazione nazionali (come il noto *American National Standardization Institute*, o ANSI).

Naturalmente una comunità di utenti particolare, nazionale o internazionale, dotata di strutture organizzative più o meno formalizzate, può decidere di sviluppare o adottare norme comuni per la rappresentazione e il trattamento dell'informazione, una sorta di standard informale a uso interno<sup>5</sup>.

In realtà molti standard nel mondo dell'informatica si impongono di fatto prima per la forza commerciale e poi vengono eventualmente adottati formalmente da un ente istituzionale; e viceversa non tutti gli standard formali riescono a scalzare uno standard di fatto.

I vantaggi di uno standard formale o informale, oltre alla portabilità sta anche nella sua apertura, ovvero nella disponibilità pubblica del-

---

<sup>4</sup> Il requisito della indipendenza dalle applicazioni sotto molti aspetti discende dalla portabilità tecnica, ma non è con essa coestensiva. Possono esistere schemi di codifica altamente portabili dal punto di vista tecnico, ma troppo orientati a una applicazione specifica, ciò che potrebbe causare problemi nella riutilizzo dei dati per altri fini di ricerca.

<sup>5</sup> Questo è avvenuto, ad esempio, per la comunità internazionale degli studiosi di informatica umanistica, che ha dato vita alle norme di codifica testuale della *Text Encoding Initiative*, di cui parleremo più avanti.

le sue specifiche. Chiunque è in grado di sviluppare applicazioni che usano i formati definiti negli standard, e nella misura in cui uno standard si consolida e non subisce mutazioni ogni applicazione conforme a esso è in grado di gestire dati codificati in base alle norme.

La pubblicità delle specifiche inoltre stimola il mercato degli applicativi, ed rende possibile lo sviluppo di software a distribuzione gratuita, una pratica molto diffusa nell'ambito della comunità scientifica legata alla rete di comunicazione Internet.

### **I tipi di schemi di codifica**

I sistemi informatici per la rappresentazione e il trattamento dell'informazione testuale sono molteplici. In massima parte, come è lecito aspettarsi, si tratta di tecnologie sviluppate in ambiti lontani dalle scienze umane. Nei prossimi paragrafi cercheremo di studiarne le caratteristiche, e di vedere come e in che misura esse si conformino alle esigenze di uno schema di codifica per la ricerca umanistica.

Abbiamo diviso queste tecnologie in alcuni gruppi tipologici. Questa proposta tassonomica rispecchia in gran parte una ormai consolidata tradizione nella letteratura specializzata (sia in ambito informatico sia in ambito informatico umanistico), formalizzata in particolare in un famoso articolo di Coombs, Renear, e De Rose [1987].

#### *La codifica dei caratteri alfanumerici*

La codifica dei caratteri rappresenta il grado zero della rappresentazione di testi su supporto digitale, poiché permette di rappresentare il contenuto linguistico di un testo. Essa costituisce la base di tutti gli ulteriori sistemi di codifica testuale, sebbene possa essere usata come schema di codifica in sé<sup>6</sup>.

Come qualsiasi altro tipo di dati, anche i caratteri vengono rappresentati all'interno di un elaboratore mediante una codifica numerica binaria. Per la precisione si stabilisce una associazione biunivoca tra gli elementi di una collezione di simboli distinti (*character repertoire*)

---

<sup>6</sup> Di fatto, esistono molti archivi testuali costituiti da testi in cosiddetto formato "puro testo", o con una colorita espressione anglosassone, *plain vanilla ASCII* [Ciotti, 1995: 173, e Calvo et al. 2003: 383].

e un insieme di codici numerici (*code set*). L'insieme risultante viene denominato tecnicamente *coded character set*<sup>7</sup>. Per ciascun *coded character set*, infine, si definisce una codifica dei caratteri (*character encoding*) che specifica l'algoritmo di mappatura tra ogni numero intero che rappresenta un carattere in un determinato *coded character set* e una o più sequenze di bit. Il numero di caratteri rappresentabili in un *coded character set* è determinato dal numero  $n$  di cifre binarie (bit) utilizzate per codificare ciascun carattere, in base alla formula  $2^n$ .

La rappresentazione di un sistema di scrittura in un *coded character set*, è assai più complessa di quanto non potrebbe apparire in prima istanza. Già la semplice rappresentazione delle varianti dell'alfabeto latino, usato nella scrittura di gran parte delle lingue occidentali, si complica per la presenza di lettere variamente qualificate (accenti, diresi, tilde, cedille, etc.). Accanto alle lettere del sistema alfabetico e alle cifre decimali, ovviamente, vanno codificati tutti i segni di punteggiatura. Se poi si considerano gli altri alfabeti mondiali, anche limitandosi a forme normalizzate e moderne, o le esigenze poste dalla trascrizione di fonti manoscritte antiche e medievali, ci si accorge che la progettazione di un sistema universale per la codifica di caratteri è assai complessa:

With the best will in the world, computer manufacturer and standard bodies alike will never be able to represent all the possible symbols occurring in written texts in a single universally agreed code set, simply because these symbols do not form a closed set. [Burnard, 1995: 43]

Questa limitazione tecnica ha determinato la creazione nel corso degli anni di molteplici *coded character set*, molti dei quali sviluppati autonomamente dai produttori di sistemi informatici. Si è insomma determinata una sorta di 'Babele informatica'.

A partire dagli anni sessanta gli enti di standardizzazione nazionali e internazionali (ISO ed ANSI in primo luogo) hanno cercato di mettere ordine nei vari sistemi di rappresentazione, elaborando e aggiornando diverse tavole di codici standard. In generale gli standard internazionali piuttosto che definire una singola pagina di codici prescri-

---

<sup>7</sup> È consuetudine rappresentare un *coded character set* in forma di tabella, in cui ogni cella, ordinata e indicizzata, contiene un simbolo (ragione per cui viene anche denominato *code page* o *code table*).

vono delle regole che governano la costruzione di sistemi di codifica nazionalizzati. A ogni standard corrisponde dunque una famiglia di *coded character set*. Naturalmente alcuni esponenti di ciascuna famiglia hanno in realtà avuto la prevalenza, specialmente nelle telecomunicazioni e nel trasferimento dati. Parte del lavoro di standardizzazione ha anche influenzato la tavole di codici implementate nei sistemi commerciali.

Il più antico standard internazionale è quello denominato *ISO 646*, che adotta una codifica a 7 bit (e dunque permette di rappresentare 128 diversi simboli). La sua definizione è iniziata nel 1963, ed è terminata nel 1967, anche se da allora ha subito diverse revisioni. Nel 1968 l'ANSI, l'ente nazionale di standardizzazione degli USA, registrò la prima versione nazionale, battezzata *American Standard Code for Information Interchange* (ASCII). Allo stato attuale questa è l'unica pagina di caratteri che viene interpretata e trasmessa univocamente su qualsiasi sistema informatico. Infatti, tutte le successive tavole di caratteri standard e commerciali ne hanno incorporato i caratteri nelle medesime posizioni.

Ben presto la capacità di rappresentazione di uno standard a 7 bit si è rivelata troppo angusta per le complesse esigenze di scambio dell'informazione. Per questo sin dagli anni settanta sono stati sviluppati diversi esemplari di codici a 8 bit (256 caratteri). Una consistente sistemazione e normalizzazione nella codifica informatica dei caratteri, tuttavia, si è avuta solo agli inizi degli anni ottanta, con la elaborazione da parte della *European Computer Manufacturer's Association* (ECMA) di una serie di *coded character set* a 8 bit, successivamente assunta dalla ISO e rilasciata come standard internazionale con il nome di *ISO 8859*.

Ciascuna tavola in questo standard è completamente definita, e contiene 256 caratteri, di cui i primi 128 coincidono con quelli della *ISO 646/ASCII*, mentre i restanti sono adattati a vari sistemi di scrittura nazionali. Tra questi *character set* il più diffuso è senza dubbio lo 8859-1, noto come *ISO Latin 1*, che contiene i caratteri necessari per scrivere in tutte le principali lingue occidentali.

Al fine di risolvere in modo definitivo – entro i limiti del possibile – il problema della rappresentazione di caratteri negli anni novanta sono stati avviati due progetti paralleli volti a specificare un *code set* universale: il primo in sede ISO ha portato alla formalizzazione dello standard *ISO 10646-1, Universal Character Set* (UCS); il secondo, promosso da un consorzio di aziende private attive nel settore infor-

matico, ha definito lo standard *Unicode*<sup>8</sup>. Fortunatamente questi due standard sono stati allineati. La caratteristica principale di Unicode e UCS è il fatto di non definire una sola codifica binaria a priori. In questo modo la definizione del repertorio di caratteri è indipendente e non limitata dal modo in cui i singoli possono essere rappresentati in memoria<sup>9</sup>. La versione più recente di Unicode (Unicode 4.2) contiene oltre 95 mila caratteri e specifica tre modalità di codifica: UTF-8, UTF-16 e UTF-32. Non possiamo in questa sede diffonderci sui dettagli tecnici. Basti dire che la codifica UTF-8, la più diffusa, adotta uno schema variabile di bit per rappresentare diversi sottoinsiemi di caratteri. In particolare i caratteri che coincidono con il set ASCII sono codificati mediante 8 bit di cui il primo è sempre posto a zero (e dunque di fatto coincidono con il codice ASCII), mentre i caratteri dei precedenti set a 8 bit (i cui codici numerici vanno da 127 a 2047) usano sempre 16 bit. In questo modo si ottengono due risultati: l'occupazione di memoria di un file in UTF-8 non cresce in modo irragionevole poiché statisticamente i caratteri del set ASCII sono di gran lunga i più presenti in gran parte dei documenti elettronici in lingue occidentali; le modifiche da apportare ai software applicativi affinché possano gestire correttamente la codifica UTF-8 sono meno onerose.

### *Le codifiche presentazionali e WYSIWYG*

Le codifiche dei caratteri sono la base di qualsiasi schema di codifica testuale, ma non esauriscono certamente i problemi di rappresentazione delle caratteristiche di un testo. Già una semplice analisi di un testo a stampa ci permette di evidenziare una serie di fenomeni editoriali che superano la semplice sequenza di simboli alfanumerici: la segmentazione del testo in macrounità, la presenza di titoli e sottotitoli, le enfasi, etc.

---

<sup>8</sup> Si veda <http://www.unicode.org>.

<sup>9</sup> In realtà un limite c'è poiché la codifica naturale di UCS/Unicode usa 32 bit di cui uno riservato, dunque il massimo numero di caratteri rappresentabile è  $2^{31}$ , un numero sufficientemente ampio da includere tutti i sistemi di scrittura storicamente conosciuti!

Nella letteratura scientifica si parla propriamente di sistemi di codifica testuale solo per i sistemi che consentono l'inclusione di informazione sulla struttura editoriale o grafica di un testo. La prima tipologia di sistemi di codifica propriamente detti che prendiamo in considerazione è quella delle codifiche presentazionali. Con codifiche presentazionale si intende qui un sistema di trattamento dei testi che utilizza dispositivi grafici e visuali per la rappresentazione di caratteristiche del testo. In sostanza si tratta di sistemi che mirano a ottenere una riproduzione sullo schermo di un elaboratore dell'aspetto grafico di una pagina stampata. Così, un segmento di testo enfatizzato viene codificato attraverso un font in stile corsivo; un titolo di capitolo con un font in stile grassetto e una determinata spaziatura dal resto del testo.

Poiché hanno lo scopo di ottenere una rappresentazione virtuale quanto più possibile mimetica del testo a stampa i sistemi di codifica presentazionali vengono (orribilmente) definiti sistemi *WYSIWYG*, acronimo dell'inglese *What You See Is What You Get*. Questo tipo di codifica viene usata dalla maggior parte dei sistemi di word processing interattivi attualmente diffusi sul mercato per le diverse piattaforme di personal computer.

La memorizzazione del file e delle codifiche avviene in maniera invisibile all'utente, che le introduce mediante interazione con vari dispositivi di interfaccia (menu, pulsanti virtuali, combinazioni di tasti), e ne vede gli effetti direttamente sul video. Questo ha prodotto l'erronea convinzione che i sistemi interattivi non usano codifiche.

Al contrario i testi prodotti con un sistema interattivo si basano su dei formati di dati molto complessi, che inseriscono comandi e caratteri di controllo all'interno del file di testo. Questo si traduce in un grosso limite alla portabilità. Tali formati, infatti, sono strettamente legati all'applicazione che li ha prodotti, o, in termini tecnici, sono "formati proprietari". Un documento realizzato con un dato sistema può essere visualizzato e modificato solo se si possiede il software con cui è stato creato. Inoltre questi sistemi sono nella maggior parte dei casi applicazioni di elaborazione testi, e dunque un documento prodotto con questi sistemi non si presta facilmente a essere sottoposto a procedimenti computazionali di analisi testuale o di reperimento dell'informazione.

*Le codifiche procedurali o orientate al documento*

Con le codifiche procedurali entriamo nel dominio proprio dei sistemi definiti markup language, linguaggi di marcatura o codifica del testo. Si parla di linguaggi perché questi sistemi di codifica si basano su un insieme di istruzioni di codifica, ciascuna dotata di particolari funzioni, e di una sintassi che specifica come e quando utilizzarle all'interno del file di testo.

In modo parallelo ai linguaggi di programmazione, anche i linguaggi di markup possono essere divisi in due tipologie: linguaggi procedurali, che nella letteratura vengono indicati anche come *specific markup language*; e linguaggi dichiarativi o descrittivi, detti anche *generic markup language*<sup>10</sup>.

Le codifiche procedurali sono state storicamente le prime a fare la loro comparsa. Sono state usate nei sistemi informatici per la composizione testuale, i cosiddetti batch formatter, sin dagli anni sessanta. Questi ricevono come input un file di dati che contiene la sequenza di caratteri del testo e inframmezzate in esso le istruzioni che specificano i diversi trattamenti di composizione da operare sul testo. Quando il programma trova una data istruzione, esegue le relative operazioni di composizione. Ad esempio stampa un spazio, o modifica il corpo del carattere. Queste codifiche vengono dette procedurali appunto perché specificano le operazioni che un certo sistema deve compiere sul testo per ottenere un determinato output (generalmente su carta). Esse sono dunque orientate alla creazione di un documento (generalmente tipografico) dotato di una certa impaginazione e impostazione grafica.

L'uso dei sistemi di codifica procedurale permette di rappresentare determinate caratteristiche grafiche del testo e dunque di separare alcune porzioni o segmenti di testo. Un titolo di capitolo ad esempio viene identificato da una serie di comandi che specificano delle spaziature verticali, una data giustificazione e un cambiamento nella dimensione o nello stile del font. Ovviamente ogni volta che un titolo ricorre la stessa sequenza di istruzioni viene applicata al testo del titolo. Tuttavia, se nel testo il titolo dell'appendice ha lo stesso formato

---

<sup>10</sup> Si vedano su questo: Herwijnen, 1994: 17; Bryan, 1987: 6; Coombs et al., 1990; Ciotti (a c. di), 2005.

grafico, non esiste alcun modo di distinguerli durante una ricerca automatica.

I sistemi di codifica procedurale sono per definizione orientati a una singola applicazione, ad esempio un sistema di composizione tipografica. Se lo stesso testo deve essere usato in un sistema automatico di generazione delle concordanze, si rende necessario una ulteriore codifica del testo, eventualmente eliminando i codici dell'applicazione di stampa che potrebbero arrecare errori nei processi computazionali del generatore di concordanze. Insomma, la portabilità di un testo codificato con sistemi procedurali è molto limitata, se non nulla. Per ovviare a queste limitazioni, e avere un sistema di codifica che fosse quanto più possibile generico, intorno agli inizi degli anni '70 venne sviluppato il concetto di un linguaggio di markup dichiarativo o descrittivo.

### *Le codifiche dichiarative o orientate al testo*

L'idea di un linguaggio di codifica dichiarativo si sviluppa come evoluzione dei sistemi di codifica procedurale. In particolare deriva dalla facoltà, offerta da alcuni di questi, di definire delle cosiddette "macro", ovvero gruppi di comandi di formattazione a cui ci si può riferire come a una unica istruzione: come dire, ad esempio, "applica la serie di comandi per la formattazione di un titolo di capitolo".

I linguaggi di markup dichiarativi assumono come primario proprio questo tipo di identificazione delle caratteristiche del testo, invertendo in un certo senso il processo: invece di specificare quali operazioni di formattazione vanno effettuate in un particolare punto del testo, si dichiara che un dato segmento testuale è istanza di un tipo di struttura editoriale del testo; insomma, si dichiara: "questo è un titolo". Un sistema di codifica dichiarativo dunque è orientato alla rappresentazione delle caratteristiche o elementi che costituiscono un testo, indipendentemente dalle finalità specifiche per le quali il testo è stato memorizzato e codificato, e dalle contingenti forme di realizzazione grafica che vengono utilizzate su un determinato supporto fisico (a stampa o su video).

Come i linguaggi procedurali, anche quelli dichiarativi vengono utilizzati inserendo all'interno del file di testo sequenze di caratteri, generalmente dette tag (etichette o marche), delimitate da caratteri speciali. Queste possono essere utilizzate indifferentemente da un sistema di composizione e stampa per applicare determinati processi di for-

mattazione, da un sistema di visualizzazione su video (browser), da un sistema di *information retrieval* strutturata per trovare tutte le stringhe di caratteri all'interno di un dato segmento codificato, da un sistema di analisi testuale (che peraltro può utilizzare le codifiche per produrre risultati molto più rilevanti dal punto di vista scientifico), e così via.

La disposizione alla rappresentazione di strutture astratte non pone limiti alla natura e tipologia delle caratteristiche testuali che si possono codificare in un testo elettronico. Piuttosto che la struttura editoriale astratta di testo si può rappresentare la sua struttura grammaticale, o la struttura retorica. In ultima analisi usando un sistema dichiarativo è possibile codificare anche la struttura fisica di un documento: ovvero, è possibile costruire uno schema di codifica presentazionale usando un linguaggio dichiarativo. Questo consente di effettuare delle vere e proprie trascrizioni diplomatiche elettroniche di un testo, mantenendo i vantaggi di uno schema di codifica dichiarativo. La potenza espressiva di un markup language dichiarativo non si limita alla tipologia di elementi che possono essere rappresentati. Se il linguaggio è dotato di una sintassi che permette di specificare le relazioni tra gli elementi, essa può essere usata per rappresentare la struttura e l'organizzazione del testo a un determinato livello di descrizione, o i rapporti tra elementi appartenenti a diversi livelli.

Queste caratteristiche logiche, fondamentali per l'applicazione nella ricerca umanistica, offrono notevoli vantaggi nel trattamento automatico dei testi. Coombs, Renear e De Rose nell'articolo menzionato precedentemente evidenziano come un sistema di codifica dichiarativa assista un autore nel processo di scrittura, poiché focalizza l'attenzione sul contenuto di un testo (o sulla struttura del contenuto) piuttosto che sulla sua forma grafica. Inoltre è fortemente agevolata l'applicazione di strumenti di *outlining*, ovvero di progettazione e di gestione dei contenuti di un testo.

I sistemi di markup dichiarativo introducono consistenti vantaggi anche nei processi produttivi editoriali e nella gestione dei flussi informativi aziendali. Poiché un medesimo schema di codifica dichiarativo può essere utilizzato in molteplici forme di trattamento informatico, i costi di produzione e gestione di una base dati testuale vengono fortemente ridotti. Ad esempio, nell'industria editoriale un testo codificato può essere facilmente sottoposto a diversi processi di formattazione senza subire modifiche, ottenendo dunque diversi formati di output su diversi media e supporti: stampa di qualità, stampa economica, edizione su supporto ottico, diffusione su rete telematica. Inol-

tre, se il linguaggio di codifica risponde a determinati criteri di standardizzazione, il testo è trasferibile su qualsiasi piattaforma disponibile senza costi di ricodificazione.

I sistemi di codifica dichiarativa peraltro si prestano ottimamente per rappresentare strutture complesse come riferimenti incrociati e collegamenti tra elementi all'interno di un testo e tra più testi. In questo modo è possibile realizzare delle cosiddette "basi di dati ipertestuali" completamente indipendenti dai particolari software applicativi che sono deputati a gestire la visualizzazione e ad attivare i collegamenti (*link*)<sup>11</sup>. Le informazioni sulla struttura dell'ipertesto infatti sono espresse dichiarativamente insieme e accanto al testo e non specificate in modo procedurale attraverso comandi e dispositivi proprietari del software.

### *Altri sistemi di codifica: il database e la rappresentazione grafica digitale*

Chiudiamo questa nostra rassegna sulle tipologie di schemi di codifica con alcune strutture dati che generalmente non vengono considerate adeguate alla rappresentazione di informazioni genericamente testuali. Parliamo dei database strutturati e dei formati grafici per la digitalizzazione di immagini. Non ci addentreremo in una discussione tecnica sui vari tipi di formati di dati grafici, o sulla struttura di un sistema di database. Ciò che ci preme evidenziare è che anche queste strutture dati possono essere utilizzate per rappresentare testi o caratteristiche testuali.

Un database è, come noto, una collezione di dati strutturata per record o schede. Ogni scheda identifica un oggetto di un dominio mediante alcune caratteristiche o tratti distintivi significativi in un contesto dato. Ad esempio il catalogo di una biblioteca è un database che descrive i libri in base al loro autore, titolo, editore, etc. Queste caratteristiche hanno senso nell'ambito di un sistema di organizzazione del

---

<sup>11</sup> Un sistema informativo che raccoglie tutti vantaggi dei sistemi di codifica dichiarativa è costituita dalla nota applicazione ipertestuale sulla rete Internet World Wide Web, che si basa su un formato di file, HTML, basato sulla sintassi del più evoluto sistema dichiarativo, lo Standard Generalized Markup Language.

patrimonio di una biblioteca e con il fine di consentire a un utente di sapere se un certo testo è parte di questo insieme.

Il classico esempio che abbiamo descritto, quello del catalogo di una biblioteca, fornisce altresì un esempio di come un sistema di database possa essere usato per rappresentare testi. Se in questo caso la rappresentazione è molto astratta, nulla vieta di immaginare altre tipologie di applicazioni [Buzzetti, 1995 e 1996]. Un testo infatti può essere visto come una sequenza di elementi omogenei discreti descrivibili sotto il rispetto di un numero di caratteristiche finito. Una sequenza di morfemi, ad esempio, individuati dalle determinazioni morfosintattiche. Si ottiene in questo modo una rappresentazione sub specie di base dati della struttura morfosintattica del testo, che può essere sottoposta a interrogazioni e ricerche.

Questo tipo di struttura dati è in effetti usata dalla maggior parte delle applicazioni di analisi testuale. Infatti un database offre dei notevoli vantaggi dal punto di vista delle prestazioni computazionali e della velocità di ricerca, anche se richiede in generale una ingente quantità di memoria per l'archiviazione. Per contro essa non offre la flessibilità e la possibilità di descrivere la struttura di un testo a diversi livelli data dai sistemi di codifica dichiarativi; infatti un database è per definizione deputato a individuare un dominio delimitato e predefinito di caratteristiche di un certo oggetto, o dominio di oggetti<sup>12</sup>.

Se la natura rappresentazionale di un sistema di database è dichiarata, non è vero il medesimo per quanto riguarda le grafiche computerizzate. Infatti al momento della percezione sul video dell'elaboratore le immagini digitali si offrono come totalità analogiche, spesso di qualità notevolmente inferiore rispetto a quella offerta dalla stampa fotografica. Tuttavia, come ogni altro tipo di informazioni, anche le immagini debbono essere trasformate in strutture dati formali per poter essere archiviate e gestite da un elaboratore. Una immagine digitale dunque non è una semplice riproduzione, come potrebbe sembrare, ma

---

<sup>12</sup> Per questo si presta ad applicazioni in campi in cui i flussi informativi sono, per quanto complessi e di grande estensione, dotati di una struttura rigida, predeterminata e non soggetta a frequenti variazioni, come le applicazioni gestionali: gestione di un magazzino, di un archivio, di un sistema di contabilità, di un sistema bancario. Già la gestione di una biblioteca, con la grande varietà di casi e di pratiche catalografiche diverse, mostra i limiti dei sistemi di gestione database relazionali o pseudo-relazionali a campi fissi, i più diffusi attualmente.

una vera e propria rappresentazione. In quanto tale esse può essere sottoposta a procedimenti di analisi computazionale [Buzzetti, 1995 e 1996].

Esistono due tipi di formati digitali per la rappresentazione di immagini: i formati vettoriali, e i formati pittorici. I formati vettoriali codificano le immagini attraverso insiemi di funzioni corrispondenti a linee e curve. Ogni volta che la figura viene visualizzata le funzioni vengono calcolate e le curve disegnate. In un certo senso la grafica vettoriale costituisce l'equivalente delle tavole di caratteri per i testi, poiché riduce una immagine a un complesso di figure primarie.

Questo formato, che ha il vantaggio di non perdere risoluzione all'aumentare delle dimensioni, e di non richiedere grandi quantità di spazio nella memoria, è usato nella creazione di immagini fortemente geometrizzate (ad esempio nel disegno architettonico, o CAD, e nella progettazione in genere), ma non si presta, almeno per ora, alla digitalizzazione di immagini realistiche.

I formati pittorici invece si basano sulla segmentazione di una immagine in una griglia cartesiana di punti, ognuno dotato di un determinato colore o tono di grigio. Tutti i sistemi di digitalizzazione di immagini usano fondamentalmente questa tecnologia. Infatti essa si presta assai facilmente ai procedimenti di scansione degli scanner e delle fotocamere digitali.

Sebbene in questo caso l'immagine non venga codificata per unità significanti, sono applicabili processi di analisi computazionali anche su questi formati. Ad esempio, nel settore della rappresentazione testuale, è possibile sezionare parti di una immagine e collegarle con segmenti testuali, per realizzare una edizione elettronica che rechi immagini dei testimoni di un'opera e le relative trascrizioni diplomatiche, accanto a un eventuale testo critico, si da offrire agli studiosi tutti gli strumenti necessari per verificare il lavoro dell'editore, ed eventualmente emendarlo, ove lo ritengano necessario [Thaller, 1992; Buzzetti, 1995 e 1996]. Allo stesso modo è possibile analizzare riproduzioni digitale di corpora di opere pittoriche, catalogando motivi, stili o elementi iconografici<sup>13</sup>.

---

<sup>13</sup> Un sistema sviluppato a questo fine è il sistema *κλειθ* sviluppato da Manfred Thaller, che permette di associare annotazioni a parti di immagini, eventualmente strutturandole in maniera gerarchica [Thaller, 1993].

## XML (Extensible Markup Language)

Il concetto di linguaggio di markup dichiarativo e descrittivo è strettamente legato allo sviluppo e all'affermazione di una particolare famiglia di linguaggi di codifica che hanno avuto e hanno ancora oggi un ruolo centrale nella gestione informatica dei testi. Ci riferiamo allo *Standard Generalized Markup Language* (SGML) e al suo successore, *Extensible Markup Language* (XML).

La storia di SGML inizia alla fine degli anni 60 nei laboratori della IBM, dove Charles Goldfarb ideò il suo diretto predecessore, battezzato *Generalized Markup Language*<sup>14</sup>. Nel 1978 l'ente statunitense per la standardizzazione, ANSI, sponsorizzò un progetto che aveva l'obiettivo di individuare o definire un possibile standard per l'archiviazione e la trasmissione di documenti digitali, alla cui direzione fu chiamato Goldfarb. All'inizio degli anni ottanta questa iniziativa venne avocata dall'ISO e nell'ottobre del 1986 SGML divenne ufficialmente uno standard internazionale.

L'accoglienza di SGML fu molto positiva, sebbene esso fosse assai complesso. Ben presto numerose comunità lo adottarono per sviluppare linguaggi di markup; ma soprattutto pochi anni dopo Tim Berners Lee, l'ideatore del World Wide Web, lo utilizzò per definire formalmente il linguaggio con cui codificare i documenti da pubblicare nel Web stesso, il notissimo *Hyper Text Markup Language* (HTML).

Proprio la necessità di potenziare la capacità di gestione dei documenti per il Web ha portato allo creazione di XML. Il suo sviluppo ha avuto inizio alla fine del 1996, nell'ambito del *Web Consortium* (W3C)<sup>15</sup>. Dopo oltre un anno di lavoro, nel febbraio del 1998 sono state pubblicate le specifiche ufficiali, disponibili sul sito del W3C all'indirizzo <http://www.w3.org/XML> (W3C 1998).

XML è un sottoinsieme di SGML, di cui eredita la sintassi e la logica di funzionamento, fatte salve alcune innovazioni che semplificano notevolmente la creazione di applicazioni software. Si tratta di un

---

<sup>14</sup> GML era alla base di un potente sistema di fotocomposizione automatica conosciuto come *Document Formatting Facility*.

<sup>15</sup> Si tratta di una organizzazione *no profit* ufficialmente deputata allo sviluppo degli standard tecnologici per il Web che raccoglie centinaia di aziende, enti, centri di ricerca e singoli specialisti coinvolti più o meno direttamente nel settore delle tecnologie Web. Per una ricostruzione storica si veda Berners-Lee 2001.

linguaggio sviluppato esplicitamente per rappresentare dati testuali complessi in formato digitale in modo indipendente da particolari sistemi e dispositivi informatici e facilitare così il trasferimento e la preservazione delle informazioni. Questi obiettivi sono stati conseguiti attraverso tre principi fondamentali:

- l'uso di un sistema di codifica dichiarativo
- il concetto di metalinguaggio di codifica
- il concetto di classe di documenti

I vantaggi offerti da un sistema di codifica dichiarativa sono stati già discussi. Resta da dire che la nozione stessa di codifica dichiarativa è stata enunciata sin dai primi anni 70 in stretta relazione con lo sviluppo dello SGML, o più precisamente del suo predecessore GML, entrambi ideati da Charles Goldfarb.

Per quanto riguarda il secondo punto va detto che XML non è propriamente un linguaggio di codifica, bensì un metalinguaggio. Infatti esso consiste di una serie di norme sintattiche astratte, attraverso le quali è possibile definire infiniti linguaggi di codifica attuali. Ognuno di questi costituisce una applicazione XML, e dunque ne eredita i caratteri di standardizzazione e generalità, ma al contempo prende forma sotto la spinta di esigenze di rappresentazione e trattamento particolari e specifiche.

La definizione di un linguaggio di codifica XML a sua volta ruota intorno al concetto di classe o tipo di documenti. Un tipo di documento è la classe di tutti i documenti che presentano le medesime caratteristiche strutturali: ad esempio il tipo di documento “lettera commerciale”, il tipo di documento “testo narrativo”, o il tipo di documento “antologia poetica”. Tale struttura astratta viene specificata individuando gli elementi logici che la costituiscono (ad esempio: capitolo, titolo, paragrafo, nota, citazione, ecc.) e le relazioni che tra questi intercorrono, relazioni che possono essere di natura gerarchica e ordinale. Infatti in XML la struttura di un documento viene descritta come un albero etichettato ordinato, che per inciso e non a caso è anche il grafo delle derivazioni di una grammatica non contestuale. Ad esempio una antologia è costituita da una sequenza di poesie. Ogni poesia può essere preceduta da un titolo. Ogni poesia è a sua volta costituita da una sequenza di versi, e così via.

La natura astratta e dichiarativa di XML, permette di estendere la nozione di tipo di documento alle differenti modalità di sezionare e analizzare un oggetto complesso come è il testo, determinate dai diversi approcci disciplinari. Un testo può essere visto come una sequenza di capitoli, paragrafi, citazioni o liste. Ma anche come una sequenza di macroenunciati, frasi, sintagmi verbali e nominali, nomi verbi e aggettivi. O ancora, come una sequenza di frontespizio, pagine recto e verso, titoli correnti, colonne di scrittura, note a margine, cancellazioni e danneggiamenti, grafemi o pittogrammi.

La struttura di una classe di documenti XML può essere specificata esplicitamente mediante la definizione di una vera e propria grammatica formale, denominata *Document Type Definition* (DTD). Il formalismo per specificare la DTD è una eredità di SGML ed è completamente stabilizzato e ampiamente adottato. Tuttavia, per ovviare ad alcuni limiti espressivi di tale formalismo, sono stati proposti numerosi formalismi alternativi, denominati *Schema Definition Language* (dove il termine 'schema' eredita il ruolo di 'tipo di documento')<sup>16</sup>.

Una DTD è costituita da un elenco di dichiarazioni che identificano gli elementi di un testo (element, in termini tecnici), e il loro modello di contenuto, ovvero l'insieme di sottoelementi e/o caratteri (l'unità minima non scomponibile) che ciascun elemento può contenere e i relativi rapporti. Questi ultimi sono essenzialmente rapporti gerarchici, come abbiamo visto. A ogni elemento inoltre possono essere associati uno o più attributi che ne specificano ulteriori caratteristiche o funzioni non strutturali: ad esempio l'aspetto fisico di un titolo o di una enfasi.

Una volta definito formalmente, un linguaggio di markup XML può essere utilizzato per creare singoli documenti, che ne dovranno rispettare i vincoli grammaticali, oltre a conformarsi alle norme generali di sintassi XML. Nel documento a ciascun elemento corrisponde una coppia di marcatori. La sintassi prevede che i marcatori siano racchiusi tra parentesi acute. Ogni elemento viene identificato da un mar-

---

<sup>16</sup> Tra le varie proposte elaborate due sembrano poter ambire al ruolo di successori della DTD: *W3C XML Schema*, il linguaggio sviluppato dal W3C (<http://www.w3.org/XML/Schema>) e *Relax NG* (<http://www.oasis-open.org/committees/relax-ng>), sviluppato da James Clark (uno dei 'guru' del settore) e supportato dall'OASIS<sup>16</sup> (*Organization for the Advancement of Structured Information Standards*)

catore iniziale e uno finale (costruito premettendo una barra al nome del marcatore iniziale), a meno che non sia un elemento vuoto (che non contiene cioè sotto-elementi o testo), nel qual caso è identificato solo da un marcatore iniziale che tuttavia presenta una barra prima del carattere '>'. Ecco un esempio di testo in XML:

```
<lg>
<l>S' i' fosse foco, ardere' il mondo;</l>
<l>S' i' fosse vento, lo tempestarei;</l>
<l>S' i' fosse acqua, i' l'annegherei;</l>
<l>S' i' fosse Dio, mandereil en profondo;</l>
<lg>
```

La conformità del documento alla DTD o allo schema può essere (e di norma viene) verificata da un programma di analisi sintattica automatica detto *parser*. Un documento XML che risponde a questa requisito viene detto *valido*. Tuttavia, a differenza di SGML, XML ammette anche l'esistenza di documenti privi di una DTD o di uno schema espliciti. Naturalmente tali documenti dovranno comunque rispettare una serie di norme sintattiche generali (che valgono cioè per ogni documento XML), dette *vincoli di buona-formazione*.

### La Text Encoding Initiative

Sin dagli anni '80 nella vasta comunità degli studiosi di scienze del testo interessati all'applicazione di metodologie informatiche, e in quella vicina degli archivisti e dei bibliotecari, si era avvertita l'esigenza di rispondere adeguatamente ai problemi di interscambiabilità delle risorse digitali determinati dalla proliferazione dei linguaggi di codifica, e di definire uno standard per la rappresentazione di testi su supporto digitale. Nel 1987 si tenne un convegno al Vassar College di Poughkeepsie (New York), nel corso del quale si decise di dare vita a un progetto di ricerca internazionale per sviluppare uno schema di codifica che mettesse ordine in questo settore. In quella occasione furono anche stilati i cosiddetti "Poughkeepsie Principles", un elenco di principi regolativi che avrebbero dovuto guidare la definizione dello schema [Ide e Sperberg-McQueen, 1995: 5].

Nel 1988, con la sponsorizzazione delle tre maggiori associazioni professionali nel campo dell'informatica umanistica e della linguistica computazionale – la *Association for Computers and the Humanities*

(ACH), la *Association for Computational Linguistics* (ACL) e la *Association for Literary and Linguistic Computing* (ALLC) – il progetto, denominato appunto *Text Encoding Initiative*<sup>17</sup>, fu di fatto avviato, con la partecipazione di studiosi provenienti da tutto il mondo. A partire dal 1989 la TEI ha sviluppato diverse versioni del suo schema di codifica, le cui specifiche provvisorie sono state pubblicate per la prima volta nel 1991 con il titolo *Guidelines for Electronic Text Encoding and Interchange, TEI P1*. Il lavoro della TEI ha poi visto molte successive revisioni che hanno portato alla pubblicazione di altri due edizioni: una nel 1992, nel quale la struttura del linguaggio è stata notevolmente rivista; e una nel 1994 la TEI P3, che ha rappresentato di fatto la prima versione definitiva dello schema. Nel 2000 i membri della TEI hanno deciso di rendere permanente il progetto, costituendo il *TEI Consortium* (<http://www.tei-c.org>), una organizzazione internazionale senza scopo di lucro fondata allo scopo di sostenere e sviluppare lo schema di codifica TEI. Il primo risultato del TEI Consortium è stata la pubblicazione nel giugno del 2002 di una nuova edizione dello schema di codifica, la TEI P4, e del relativo manuale [Burnard, Sperberg-McQueen (a c. di), 2002]. L'aspetto più importante di questa nuova versione è la sua piena conformità con XML. Attualmente è in fase di definizione la TEI P5 – il cui rilascio è previsto per la metà del 2007 – che introdurrà una grande quantità di innovazioni.

Nella lo sviluppo delle norme di codifica, sono state perseguite due finalità fondamentali:

The goals of the TEI project initially had a dual focus: being concerned with both what textual features should be encoded (i.e. made explicit) in an electronic text, and how that encoding should be represented for loss-free, platform-independent, interchange. [Burnard, 1995: par. 2]

Il conseguimento di questi obiettivi ha determinato alcune scelte teoriche e tecniche, delle quali la principale è stata senza dubbio l'individuazione dello Standard Generalized Markup Language come linguaggio base per la definizione dello schema di codifica, cui in anni recenti è stato affiancato e poi sostituito XML. Questo ha permesso la formulazione di uno schema potente ma al contempo flessibile, in

---

<sup>17</sup> Sulla TEI, oltre alla vasta documentazione disponibile in rete, rimandiamo a Ciotti (a c. di), 2005.

grado cioè di incontrare bisogni ed esigenze di rappresentazione differenziate, e portabile.

Lo schema di codifica predisposto dalla Text Encoding Initiative è sostanzialmente una complessa, e per molti versi originale, applicazione SGML. Essa consta di una vasta Document Type Definition, in cui la sintassi SGML viene usata per specificare quali elementi e strutture testuali possono/devono essere codificati in un testo elettronico e come tale codifica deve essere rappresentata per conseguire la massima generalità e portabilità. Alla DTD si affianca un monumentale manuale di specifiche che forniscono la semantica dettagliata per tutti gli elementi definiti nella DTD (oltre quattrocento).

Naturalmente la ricerca scientifica, se da un lato si giova della disponibilità di schemi di riferimento oggettivi e condivisi per la rappresentazione dei suoi oggetti di studio, e delle analisi su di essi effettuate, non può essere completamente vincolata da sistemi che rischiano di ipostatizzare metodologie e impostazioni disciplinari particolari. Peraltro abbiamo osservato come nell'ambito delle scienze umane possono darsi numerose definizioni del concetto di testo, spesso contraddittorie tra loro. Poiché si propone di fornire una metodologia standard di codifica testuale per il complesso delle scienze umane, la TEI ha dovuto temperare le esigenze di formalità e di rigore dello SGML, con le molteplici e diverse istanze emergenti dai vari settori della ricerca umanistica.

Questo risultato è stato ottenuto attraverso una architettura modulare della DTD, che permette al responsabile della codifica di un testo di adattare lo schema alle sue esigenze specifiche di ricerca. La DTD della TEI, infatti, non si limita a definire un unico e singolo tipo di documento, ma piuttosto una serie di insiemi di elementi specializzati, divisi in quattro gruppi:

- insiemi di elementi generici eventualmente rintracciabili in ogni tipo di testo (*core tag set*);
- insiemi elementi specifici per la rappresentazione dei generi testuali fondamentali: testo in prosa, testo in versi, testo drammatico, dizionari, trascrizioni di fonti orali (*base tag sets*);
- insiemi di elementi addizionali per la rappresentazione di caratteristiche evidenziate da particolari prospettive analitiche e applicazioni specializzate: codifica di fonti primarie e di apparati di varianti, codifica di strutture morfosintattiche, rappre-

sentazione di strutture interpretative profonde, rappresentazione di strutture ipertestuali (*additional tag sets*);

Ogni studioso può costruire una propria “vista” del DTD combinando l’insieme elementi fondamentale con un solo insieme di elementi per il tipo di testo e quanti insiemi addizionali desidera.

Accanto a questa architettura modulare la TEI ha cercato di fornire strutture di codifica quanto più possibile “politeoretiche”, ovvero capaci di rappresentare le caratteristiche testuali a un dato livello secondo differenti impostazioni teoriche e metodologiche. Ad esempio la codifica della struttura morfosintattica non impone l’adesione a una visione teorica, sia essa quella “generativo-trasformativa” o quella “funzionalista”, poiché è abbastanza plastica da consentire analisi dettagliate secondo entrambe le metodologie, ed eventualmente per specificarne di nuove.

Un ulteriore carattere distintivo dello schema TEI è il tentativo di superare alcuni limiti rappresentazionali derivanti dalla struttura gerarchica intrinseca nella sintassi SGML/XML. In un testo, infatti, non è rara la compresenza di elementi appartenenti a gerarchie sovrapposte, o di strutture non circoscrivibili affatto in alcuna struttura gerarchica, che sono difficilmente rappresentabili attraverso la rigida struttura sintattica dello standard. La TEI fornisce alcuni costrutti che, pur rimanendo sintatticamente validi, permettono la rappresentazione di strutture non gerarchiche. Tuttavia la loro applicazione comporta serie difficoltà, e accentua la prolissità della codifica.

Nonostante i limiti, che potranno essere superati da future revisioni dello standard, lo schema di codifica della TEI raccoglie il frutto di oltre venti anni di studi testuali con metodologie informatiche. Esso attualmente si pone come inevitabile punto di riferimento e di confronto per tutti i progetti di ricerca che prevedano la codifica e il trattamento informatico di testo in ambito umanistico.

Piuttosto che trarre conclusioni, pertanto, preferiamo indicare alcune prospettive aperte dai temi trattati. In primo luogo la disponibilità di strumenti come la TEI, progettati e sviluppati interamente da umanisti, offrono oggi un valido supporto alla creazione di archivi digitali di testi, e consentono di passare dalla fase puramente speculativa sui nuovi media e sulle biblioteche virtuali, alla realizzazione di progetti concreti. Essi, inoltre, propongono una nuova forma di fruizione multimediale, per usare un termine alla moda, che potrebbe estendere e diversificare i soggetti della fruizione, e segnare una inversione di

tendenza in una società in cui l'accesso ai testi e alla scrittura è viepiù ridotto.

In secondo luogo la riflessione sui problemi della codifica informatica dei testi, ha aperto il campo a una intensa discussione teorica sul concetto di testo e più in generale sulla natura epistemologica delle scienze umane. L'informatica umanistica dunque è in grado di fornire gli strumenti per affrontare in modo nuovo alcuni temi fondanti e costitutivi delle scienze umane, e di contribuire sostanzialmente alla evoluzione del sapere umanistico in generale.

Infine, lo sviluppo dei sistemi di codifica è un esempio concreto che dimostra come la comunità umanistica possa dare un contributo decisivo all'evoluzione delle tecnologie dell'informazione e della comunicazione, non ristretto nei limiti specialistici, per quanto importanti, della ricerca. Questo naturalmente costringe il mondo umanistico (specialmente quello accademico), a ripensare se stesso, ma apre anche importanti prospettive a una cultura che corre il rischio di perdere i rapporti con il resto della società.

## Aspetti teorici e metateorici della codifica testuale

Nel capitolo precedente abbiamo affrontato gli aspetti squisitamente tecnici relativi ai sistemi di codifica testuale attualmente esistenti, e indicato una serie di requisiti che i sistemi di codifica dovrebbero avere per rispondere alle esigenze della ricerca in ambito umanistico. In questo cercheremo invece di analizzare alcuni problemi teorici implicati dalla codifica informatica dei testi.

Lo sviluppo e l'applicazione di sistemi di codifica infatti, è stata affiancata da una vivace e interessante discussione teorica. Sebbene nella maggior parte dei casi sia stata mossa da considerazioni pratiche, la teoria sulla codifica testuale ha spesso toccato problemi che attengono ai fondamenti e alla epistemologia delle scienze umane.

La codifica infatti è un processo assai più complesso delle semplice e meccanica correlazione biunivoca di strutture rappresentazionali. Essa impone una approfondita riflessione sul "problema del testo", per usare una felice formula di Bachtin, ovvero la determinazione di «cosa sia un testo» e di quale legame sussista tra questa determinazione, i processi dell'interpretazione e i linguaggi con i quali essa viene enunciata.

Come ricordavamo in apertura, lo sviluppo e l'applicazione dei vari schemi di codifica che sono stati di volta in volta utilizzati per il trattamento informatico dei testi, ha dato luogo a un interessante dibattito teorico. Parte di questa produzione teorica si colloca in un ambito strettamente informatico, più in particolare nel settore del document processing, e riguarda gli aspetti tecnici dei sistemi di codifica e trattamento dei testi che hanno portato allo sviluppo di SGML e XML.

La progressiva diffusione di metodologie e applicazioni computazionali in ambito umanistico e letterario, tuttavia, ha indotto molti studiosi di questo dominio disciplinare ad affrontare i fondamenti teorici ed epistemologici di tali nuove pratiche di ricerca.

Non è nostra intenzione fornire un quadro completo della produzione teorica concernente la codifica testuale, né formulare una teoria compiuta e sistematica. Piuttosto cercheremo di tematizzare i nodi fondamentali intorno ai quali si è articolata la riflessione sulla codifica testuale in ambito umanistico, con particolare attenzione ai problemi

teorici sollevati dalla applicazione di schemi basati su SGML e XML, come le norme prodotte dalla TEI.

In linea generale questa produzione teorica si è concentrata su due aspetti problematici: uno di carattere epistemologico, riguarda la natura della codifica come processo di rappresentazione.

Il secondo è piuttosto di carattere ontologico, e concerne il concetto generale di testo che «emerge» dalle teorie dei sistemi di codifica.

In realtà la distinzione tra questi due poli teoretici ha una valenza espositiva. Infatti l'analisi del processo di scoperta e rappresentazione dell'oggetto testo e della sua struttura non può prescindere da una determinazione di cosa si intenda per testo, o quantomeno da una assunzione implicita sulla sua natura; d'altra parte molte concettualizzazioni sulla natura del testo assegnano al soggetto della codifica, e alla sua competenza teorica e teorie generali sul mondo, un ruolo molto importante, se non costitutivo. Su questo, tuttavia, è opportuno ricordare quanto affermano Renear, Mylonas e Durand:

Of course, it may turn in the end that the epistemology and metaphysics of text objects are profoundly entangled with each other – but this is a conclusion one should reach as the results of argument and analysis, and not assume at the outset of an investigation of the principles implicit in current encoding practices. [Renear et al., 1993]

### **La natura epistemologica della codifica testuale**

Il primo problema teorico implicato dalla codifica informatica del testo riguarda la natura e le condizioni del processo di codifica in quanto tale. La questione si riassume nella controversa domanda se la codifica sia un processo interpretativo o semplicemente riproduttivo.

Le argomentazioni intorno a questo tema hanno un forte legame con le pratiche concrete di codifica dei testi. Infatti esse si incrociano con la polemica che vede contrapposti i fautori dei sistemi di codifica procedurale a quelli dei sistemi dichiarativi, e con le critiche rivolte allo schema di codifica SGML/TEI e ai suoi presupposti teorici.

Secondo tali critiche lo schema di codifica TEI impone al responsabile della codifica di effettuare delle scelte teoriche e interpretative che non sono pertinenti alla sua opera di semplice trascrittore. Espo-

nenti autorevoli di questa corrente sono Ian Lancashire, dell'Università di Toronto, e Mark Olsen, direttore del progetto ARTFL<sup>1</sup>. In diversi articoli e interventi nel forum telematico *Humanist*, i due studiosi hanno ripetutamente sostenuto che nella codifica è bene limitarsi a registrare esclusivamente la forma fisica del supporto e le iscrizioni in esso contenute, ciò che SGML, in quanto sistema dichiarativo, non consente.

Questa posizione, peraltro condivisa da altri studiosi, si basa su due assunti. Un assunto di tipo teorico e metodologico, in base al quale la codifica informatica del testo è una semplice trascrizione meccanica, e in quanto tale non teoricamente problematica. Un assunto di tipo tecnico, che sostiene la inadeguatezza della codifica dichiarativa ad assolvere correttamente tale compito.

### *Ancora sulla tipologia dei linguaggi di codifica*

L'argomento tecnico, per quanto non privo di fondamento, può essere sostanzialmente confutato. Si può infatti osservare come non sia vero che i sistemi di codifica dichiarativi non consentano di realizzare descrizioni della forma fisica di un supporto testuale. La distinzione tra sistemi procedurali e sistemi dichiarativi non è parallela ma ortogonale a quella tra codifica presentazionale e codifica analitica o strutturale.

Abbiamo vista sopra come uno dei vantaggi dei sistemi dichiarativi sia proprio la capacità di rappresentare molteplici strutture testuali. A maggior ragione SGML, che è un metalinguaggio dichiarativo, può essere usato per descrivere anche la forma fisica di un documento testuale. Basta, a tale scopo, sviluppare una adeguata Document Type Definition in cui siano dichiarati elementi come «manoscritto», «recto», «verso», «colonna», «rigo» etc.

Naturalmente questo è possibile se tale descrizione del supporto fisico di un testo è riducibile a una struttura gerarchica. I problemi e le difficoltà determinati dagli schemi SGML per una codifica presentazionale in effetti, sono determinati proprio da questa metastruttura

---

<sup>1</sup> Il progetto ARTFL è un archivio testuale della letteratura francese, uno dei maggiori data base testuali attualmente disponibili, sviluppato dalla Università di Chicago, <http://humanities.uchicago.edu/ARTFL/ARTFL.html>.

profonda del linguaggio (peraltro non limitati alla codifica di fonti primarie), e non dalla sua natura dichiarativa. Per questo è necessario indagare più a fondo la natura della codifica e dell'idea di testo che la codifica presuppone.

*La codifica come processo interpretativo*

Certamente di maggiore interesse è l'argomento teorico e metodologico, che sostiene la necessità di attuare una codifica del testo su supporto informatico esclusivamente meccanica e riproduttiva. Si può infatti veramente sostenere che la codifica sia un semplice processo di trascrizione? E, anche se la codifica fosse limitata alla semplice trascrizione, è possibile sostenere che questa sia un processo meccanico, completamente privo di fattori interpretativi?

La discussione svolta in apertura ci consente di enunciare la seguente definizione: la codifica testuale è la rappresentazione formale di un testo e delle sue caratteristiche mediante un linguaggio informatico. Il termine «rappresentazione» di questa definizione è molto forte: la rappresentazione informatica è un processo semiotico come molti studiosi hanno già osservato [Gardin, 1991: 60; Gigliozzi, 2003 e Orlandi, 1996]. Ogni atto rappresentazionale o semiotico implica dei processi interpretativi (inferenziali o abduktivivi, pur se di diversa complessità). Conseguentemente sosteniamo che ogni processo di codifica (inclusi quelli di cui ci occupiamo in questa sede) è il risultato di una interpretazione.

Per argomentare questa affermazione prendiamo a titolo di esempio un processo di rappresentazione apparentemente privo di interferenze interpretative: la trascrizione della sequenza di caratteri.

Il riconoscimento di un carattere su un testo scritto implica l'assunzione che una data traccia grafica «A» sia un token di un data classe astratta di tracce che identifichiamo come il carattere «a». Il riconoscimento dei caratteri di un testo dunque è una abduzione ipercodificata, nella terminologia di Eco. Se nella maggioranza dei casi chiunque è in grado di effettuare in modo automatico questa abduzione, si danno molti casi in cui questo non è facile, o è del tutto impossibile, e vengono richiesti ulteriori sforzi interpretativi. Come rileva ancora Eco, «riconoscere un fenomeno dato come il token di un dato type presuppone alcune ipotesi sul contesto espressivo e sul co-testo discorsivo» [Eco, 1990: 237].

Prendiamo ad esempio una traccia somigliante a una «a» ma mancante di una parte del piede: posso assumere che si sia verificata una abrasione sulla carta; questa assunzione può essere validata dal contesto verbale: se le lettere che seguono sono «more» sono autorizzato a inferire la presenza della parola «amore». Tuttavia potrebbe valere anche una interpretazione che individui uno spazio tra la mia iscrizione misteriosa e la sequenza more, che farebbe diventare la mia scritta «o more». Casi come questi sono tipici nella pratica filologica. Si noti che in molti casi nemmeno allargando il contesto si potrebbe arrivare a una disambiguazione certa.

Insomma per riconoscere un carattere abbiamo dovuto ricorrere a diversi livelli di competenza semiotica. In taluni casi il riconoscimento dei caratteri su un supporto scritto comporta la costruzione vera e propria di nuovi codici (ad esempio se si stanno decifrando delle iscrizioni sconosciute). Questa consapevolezza è ben presente nel lavoro di filologi e codicologi. Essi non accetterebbero in linea di principio nemmeno l'idea che è sempre possibile predicare in modo non ambiguo l'appartenenza di una certa «traccia» su un supporto testuale fisico a una certa classe di iscrizioni che chiamiamo «carattere». La trascrizione di un manoscritto dunque è un atto fortemente interpretativo. In questa nostra discussione non abbiamo fatto riferimento peraltro alla utilizzazione dei simboli del linguaggio informatico in funzione di *representamen* dei caratteri alfanumerici del testo. Questa correlazione, infatti, comporta non pochi problemi, a causa della natura sintatticamente rigida dei primi.

In ultima analisi, la codifica informatica di un testo può essere vista come il prodotto di un insieme di inferenze che vengono espresse mediante un linguaggio formalizzato. Tali processi interpretativi possono collocarsi su livelli diversi di descrizione. Si può ad esempio asserire che in un dato documento testuale una data sequenza «x» va intesa come «titolo di un capitolo». E si può inferire che una data porzione testuale è ascrivibile alla funzione narratologica «complicazione». Ogni interpretazione può godere di diversi gradi di certezza e di soggettività. In ogni caso non esiste nessun genere di rappresentazione di un testo che si possa definire libera da processi interpretativi.

La codifica informatica di un testo provvede un sistema linguistico formalizzato che permette a uno studioso di «rendere esplicita una interpretazione di un testo» [Burnard, 1995: 43], e le varie operazioni inferenziali implicite che la hanno prodotta. In questo senso i sistemi dichiarativi forniscono un potente dispositivo metalinguistico per e-

sprimere interpretazioni sul testo in modo efficiente, intersoggettivo e verificabile.

### **Teoria del testo e struttura dei linguaggi di codifica**

La codifica permette allo studioso di esplicitare le sue ipotesi interpretative. Mutuando la terminologia della filosofia della scienza, si può dire che lo schema di codifica è un linguaggio teorico usato per costruire teorie o modelli di fenomeni testuali.

L'utilizzazione di un certo linguaggio teorico implica sempre una presa di posizione ontologica sulla porzione di mondo di cui si parla. Più radicalmente, la stessa "costruzione" di un linguaggio teorico riflette un determinato modello del mondo.

La nozione di funzione modellizzante di un sistema rappresentazionale è stata formulata con grande chiarezza dal semiologo Juri Lotman. In particolare, consideriamo il seguente illuminante esempio, riferito al linguaggio simbolico della chimica:

Supponiamo ora di suddividere tutto lo spazio designato i gruppi. Per esempio, descriveremo tutto l'insieme del contenuto con l'aiuto di una lingua che abbia solo due nomi: metalli e non metalli, oppure introdurremo altri sistemi di trascrizione, finché non arriveremo alla articolazione in elementi e loro designazione per mezzo di singole lettere. È chiaro che ogni sistema di trascrizione rifletterà una certa concezione scientifica della classificazione del designato. In tal modo, ogni sistema della lingua chimica è nello stesso tempo anche un modello di una certa realtà chimica. Siamo giunti a una conclusione sostanziale: ogni lingua è non solo un sistema comunicativo, ma anche un sistema di simulazione [i. e. «modellizzazione» n. d. A.]; più precisamente entrambe le funzioni sono indissolubilmente legate. [Lotman, 1990:20]

La costruzione di un sistema linguistico determina la pertinentizzazione dell'universo attraverso le unità espressive del codice: essa sottende le operazioni di selezione e classificazione degli elementi rilevanti, le quali, se non sono indirizzate da una teoria preesistente, ne istituiscono una. Ne consegue che la struttura o forma logica del linguaggio istituisce un rapporto con la struttura del mondo.

Applicando queste osservazioni al tema del nostro discorso, possiamo dire che il linguaggio di codifica informatica del testo, implica una teoria ontologica del testo, e che la struttura formale del linguaggio di codifica implica una determinata concezione sulla struttura del testo.

Arriviamo così al secondo nodo problematico che costituisce il campo di una teoria della codifica, espresso dalla domanda: quale concezione o modello ontologico del testo è implicata nella rappresentazione informatica?

### *La teoria gerarchica degli oggetti testuali*

La definizione di assunti metateorici sulla natura del testo, come rileva Allen Renear, che ne è stato uno dei più acuti studiosi, oltre che uno degli artefici, è strettamente collegata allo sviluppo, e soprattutto alla diffusione dei sistemi di markup dichiarativi in ambito umanistico, e in particolare di SGML.

La prima generalizzazione dei fondamenti strutturali di questo particolare linguaggio è stata enunciata in un celebre articolo dall'inequivocabile titolo «What is Text, Really». La tesi formulata in questo scritto è che il testo è, in un senso importante, una “ordered hierarchy of content objects (OHCO)”, una gerarchia ordinata di oggetti di contenuto [DeRose et al., 1990:3].

Gli oggetti di contenuto testuale a cui si fa riferimento in questa teoria sono sostanzialmente le strutture editoriali astratte di cui si compone un testo: capitoli, paragrafi, citazioni, enfasi, o poesie, strofe, versi, etc. Essi sono gerarchici poiché alcuni degli oggetti testuali contengono altri, e ordinati in quanto esiste una relazione lineare tra due oggetti posti sul medesimo livello gerarchico.

La specificazione tipologica degli oggetti è determinata dal tipo di documento, che come abbiamo visto è una nozione tecnica SGML. In termini più vicini ai discorsi teorici sui testi di ambito umanistico, il genere determina gli elementi che costituiscono il testo. Reciprocamente un genere testuale è individuato dalla classe di oggetti di contenuto che contiene: il testo poetico è tale perché si articola in sequenze di elementi che chiamiamo «versi», le quali, per inciso, non hanno assolutamente nulla a che fare con la struttura verbale del testo. Se volessimo trascrivere un poema dovremmo essere in grado di segnalare questa struttura. Il medesimo discorso potrebbe essere fatto per un testo drammatico: la distinzione tra battute e didascalie è essenziale per comprendere il testo. Alcune caratteristiche sono comuni a tutti o a molti di questi tipi di testi, mentre altre sono assolutamente specifiche.

Chiaramente questa teoria del testo è fortemente determinata dalla sintassi dello SGML, che descrive appunto un documento come una struttura gerarchica. In effetti essa trova fondamento in gran parte del-

la letteratura tecnica relativa al linguaggio, dove formulazioni analoghe, anche se meno compromesse dal punto di vista ontologico, possono essere rintracciate [Goldfarb, 1981 e 1990, Coombs et al., 1990]. La teoria gerarchica, tuttavia, nel riprendere le elaborazioni di natura funzionale della letteratura tecnica le trasforma in un asserto essenzialista circa la natura del testo.

La giustificazione della teoria gerarchica del testo, che emerge dalla letteratura che la supporta, si fonda su una serie di argomentazioni pragmatiche e teoretiche, così riassunte da Renear:

1. explains the success of certain representational strategies
2. is implied by our theorizing about literature
3. matches our intuitions about what is essential and what accidental about textual identity [Renear et al., 1995]

La prima serie di argomentazioni fa riferimento all'efficienza computazionale che la struttura gerarchica mostra nelle applicazioni informatiche. In sostanza si sostiene che, poiché descrivere un testo come una struttura gerarchica ne facilita e potenzia il trattamento informatico, allora è vero che il testo è una struttura gerarchica. Queste argomentazioni coincidono spesso con l'esposizione dei vantaggi tecnici dei sistemi di markup descrittivi rispetto altri sistemi di trattamento dei testi.

La seconda argomentazione si basa sull'osservazione che ogni discorso teoretico relativo ai testi usa nelle sue spiegazioni termini come capitolo, paragrafo, verso, titolo: dunque questi oggetti testuali, che vengono nominati in teorie diverse, a volte contrastanti, debbono esistere indipendentemente dalle teorie stesse.

La terza argomentazione invece mutua una classica strategia argomentativa di tipo metafisico sulla identità: poiché noi siamo indotti ad affermare che un testo è quel testo solo se ritroviamo la stessa sequenza di parole, capitoli, titoli, indipendentemente dalle sue realizzazioni fisiche, allora tali oggetti sono ciò che costituisce la sua essenza.

Questa argomentazione può trovare fondamento in alcune interessanti analisi di Nelson Goodman (la cui filosofia ha peraltro una impostazione nominalista) sulla natura dei linguaggi della comunicazione artistica, dove egli rileva assegna il carattere «allografico» del testo letterario. Per il filosofo inglese la possibilità di stabilire l'identità di un testo è determinata dalle caratteristiche formali dei simboli che

vengono utilizzati per produrre e riprodurre il testo, i caratteri della scrittura alfabetica:

Per identificare un esemplare dell'opera o per produrre un nuovo esemplare basta verificarne la compitazione o compitare correttamente. Il fatto che un'opera letteraria è composta in una notazione definita, che consiste di certi segni o caratteri che debbono essere combinati in successione, fornisce in effetti il mezzo per distinguere le proprietà costitutive dell'opera da tutte quelle contingenti – cioè a dire, per fissare i tratti necessari e, per ciascuno di essi, i limiti di variazione ammissibili. [Goodman, 1976:102]

Un simile argomento può essere generalizzato senza alcun problema alla nozione di oggetti testuali espressa nella teoria gerarchica, di cui i caratteri del testo sono una istanza particolare.

In conclusione secondo questa teoria un testo è una struttura gerarchica di oggetti logici, e la codifica non fa altro che esplicitare questa sua struttura essenziale. I problemi di rappresentazione, in questo contesto, sono sempre dovuti a una cattiva interpretazione del testo.

### *I limiti della teoria gerarchica*

La teoria gerarchica degli oggetti testuali ha accompagnato e sostenuto l'affermazione dei sistemi di codifica dichiarativi, e in particolare l'affermazione dello SGML come linguaggio universale per la rappresentazione digitale dei testi. Su questo impianto teorico si è basata, ad esempio, la prima fase del lavoro della Text Encoding Initiative.

Tuttavia proprio l'applicazione in sede umanistica di questi schemi di codifica ha evidenziato una serie di problemi di rappresentazione che costituiscono dei veri e propri controesempi di una teoria rigidamente gerarchica degli oggetti testuali.

In primo luogo la definizione della nozione di genere si rivela estremamente sfuggente. Ad esempio alcuni testi possono essere considerati allo stesso tempo esemplari del tipo «testo in versi» e di quello «testo drammatico». E, cosa ancora più notevole, ci sono moltissimi casi in cui non esiste assolutamente accordo tra gli specialisti dei testi nell'asserire l'appartenenza di un dato testo a un tipo piuttosto che a un altro: in realtà non esiste nessun accordo definitivo tra gli specialisti su cosa sia un genere testuale, e su quanti generi testuali esistano. In effetti ci accorgiamo che la definizione di genere testuale è dipendente dal punto di vista teorico che si assume sulla nozione di genere.

L'introduzione del concetto di dipendenza dal punto di vista ha delle conseguenze veramente rilevanti. Infatti se esso è determinate nella definizione dei generi di testo e dunque delle caratteristiche essenziali di un testo che appartiene a un determinato genere, non esistono ragioni teoriche per non applicarlo alla determinazione delle caratteristiche di un singolo tipo di testo, e in ultima analisi di quelle di un singolo testo tout-court. È sufficiente ad esempio chiedere a un linguista a un filologo e a un narratologo cosa intenda per testo, o quale siano le caratteristiche che gli interessano in un testo: avremmo sicuramente una serie di risposte diverse, in cui compaiono tre insiemi di termini teorici che si intersecano solo in alcuni elementi.

Il filologo, ad esempio, considererebbe rilevante ogni minima variazione di carattere presentata da diversi esemplari materiali di un'opera (testimoni), poiché lo scopo del suo lavoro consiste nel fare ipotesi circa la compitazione originale di un testo a partire da un certo numero di testimoni che presentano «variazioni». Al contrario il narratologo è interessato a studiare come, partendo da una certa sequenza di morfemi si costruiscano quelle strutture astratte che vengono denominate «intreccio», «fabula», «strutture attanziali», «funzioni narrative». Ci accorgiamo dunque che la pratica comune nelle discipline che studiano i testi è quella di definire il loro oggetto a partire da un punto di vista interno alla disciplina stessa. Questo a sua volta si esplicita in un determinato linguaggio teorico, o nei termini di una teoria della codifica, dà luogo alla individuazione di diversi unità o elementi di contenuto testuale.

Non solo. È facile dimostrare che questi diversi insiemi di elementi di contenuto non possono essere ricondotti a una struttura gerarchica unitaria. La codifica di testi letterari, infatti, rivela spesso l'esistenza di classi di elementi testuali che si sovrappongono rompendo i confini della struttura gerarchica di un documento in formato SGML/XML [Barnard et al, 1988]. Un esempio classico di elementi strutturali che presentano questo comportamento sono la struttura metrica e quella morfosintattica di un testo poetico. Questi elementi si comportano come se appartenessero a diverse gerarchie di oggetti testuali che si sovrappongono.

### *La revisione pluralista teoria gerarchica*

Poiché una definizione «essenziale» unitaria dell'oggetto «testo» incorre in controesempi specifici e in confutazioni complessive occor-

re abbandonare l'idea platonica di struttura essenziale del testo, per privilegiare piuttosto un approccio pluralista, che preveda la coesistenza di molteplici visioni o modelli del testo, ognuno corrispondente a un particolare linguaggio teorico, nell'ambito dei quali possono emergere diverse interpretazioni locali. Questa osservazione tuttavia non implica l'abbandono della metastruttura gerarchica del modello. In particolare Renear Milonas e Durand, hanno proposto una revisione della teoria essenzialista gerarchica [Renear et al, 1995] che si basa sulla nozione di prospettiva analitica e che fa salvo il concetto di gerarchia ordinata. La loro tesi afferma che "ogni prospettiva analitica su un testo determina una struttura gerarchica di oggetti di contenuto".

Laddove "prospettiva analitica" va intesa come "una famiglia naturale di metodologia, teoria e pratica analitica". Un corollario operativo di questa tesi è che se due elementi a e b si sovrappongono, allora essi appartengono a due diverse prospettive analitiche.

Si noti che il concetto di prospettiva analitica ha un parziale corrispettivo teorico in molte teorie strutturali del testo letterario e nella linguistica testuale, dove compare spesso la nozione di livello testuale [Segre, 1984].

La revisione pluralista della teoria gerarchica si è rivelata sufficientemente esauriente per fondare una grande parte delle applicazioni di schemi di codifica SGML e XML nell'ambito della ricerca umanistica. La stessa idea di compresenza di gerarchie multiple è prevista tra i costrutti sintattici di SGML<sup>2</sup>.

Una implementazione avanzata ed elegante di questa teoria è alla base dello schema di codifica definitivo sviluppato dalla Text Encoding Initiative. Esso infatti ha una struttura modulare, in cui ogni modulo corrisponde alla rappresentazione di un determinato punto di vista metodologico sul testo.

---

<sup>2</sup> Si tratta della caratteristica opzionale *CONCUR*, che permette di associare un medesimo documento a più DTD. Essa tuttavia viene scarsamente implementata nei software di gestione SGML attualmente disponibili a causa di alcune inconsistenze sintattiche che in taluni casi può generare. XML non è dotato di questo costrutto.

*Il pluralismo metodologico e la nozione di sottoprospektiva*

La estensione pluralista della concezione gerarchica del testo ha rappresentato il fondamento teorico di gran parte delle correnti pratiche di codifica informatica dei testi. Tuttavia alcune recenti applicazioni hanno portato alla individuazione di alcuni controesempi pratici e alla contestazione dei suoi assunti teorici.

La più rilevante critica di carattere teorico riguarda la insufficiente determinazione del concetto di prospettiva analitica, che è una nozione decisamente informale. Poiché a essa corrisponde, dal punto di vista della rappresentazione informatica, la sovrapposizione di gerarchie testuali, fenomeno formalmente individuabile, si rischia di approdare a una definizione circolare di prospettive analitiche e gerarchie sovrapposte. Insomma, come dire: una prospettiva analitica si individua in presenza di un fenomeno di sovrapposizione di elementi, e un fenomeno di sovrapposizione è determinato dalla presenza di due prospettive analitiche.

Ma è vero che a ogni coppia di elementi che si sovrappongono corrispondono due distinte prospettive teoriche? Forti dubbi circa la veridicità di questa asserzione emergono dalla occorrenza di alcuni oggetti testuali che, pur appartenendo ragionevolmente a una medesima prospettiva analitica si sovrappongono; o ancora fenomeni di interrelazione tra diverse gerarchie, che corrispondono a termini dei linguaggi teorici “tradizionali” sui testi. Ad esempio, nello studio dei testi poetici esiste un termine teorico, *enjambement*, che ha come referente un fenomeno di intersezione tra due gerarchie, la gerarchia della struttura linguistica e quella della struttura prosodica.

Al fine di giustificare l'occorrenza di gerarchie sovrapposte all'interno di una medesima prospettiva analitica, è stata proposta da Renear [1995] e dai suoi collaboratori una terza revisione della teoria gerarchica che, introduce la nozione di sottoprospektiva, e afferma che se due oggetti testuali evidenziati da una prospettiva teorica si sovrappongono, allora essi appartengono rispettivamente a due sottoprospektive diverse della prospettiva teorica principale<sup>3</sup>. Una sottoprospektiva

---

<sup>3</sup> In questo senso l'*enjambement* crea una sovrapposizione poiché è un fenomeno che identifica due sottoprospektive: quella individuata dalla “analisi metrica” e quella individuata dalla “analisi sintattica” nell'abito dell'analisi del testo poetico.

può essere informalmente intesa come una corrente metodologica o una specializzazione di una famiglia disciplinare, dove i metodi della seconda sono ereditati dalla prima ma non viceversa.

Questa ultima revisione della teoria gerarchica abbandona qualsiasi assunto di tipo essenzialista. In questa ottica il testo diventa un sistema a più livelli, che corrispondono a diversi punti di vista dell'osservatore.

Si deve rilevare che la complessità rappresentazionale soggiacente a questa teoria, portata alle sue estreme conseguenze, eccede i limiti sintattici di un formalismo di codifica come XML (nella sua specificazione attuale). Lo standard, infatti, non è dotato di costrutti sintattici adeguati alla rappresentazione di molteplici sottoprospective gerarchiche concorrenti che si sovrappongono ma che possono anche collegarsi e interrelarsi.

Tuttavia vengono mantenuti ancora i due presupposti fondamentali che erano validi nella primitiva versione gerarchica, e che giustificano in ultima analisi l'applicazione di un metalinguaggio di codifica come SGML:

- il testo è un oggetto reale dotato di una sua struttura che corrisponde alla struttura del linguaggio di rappresentazione;
- la struttura o meglio le strutture del testo sono strutture gerarchiche: le sottoprospective, comunque esse siano definite, danno luogo a gerarchie.

L'impianto finale dello schema di codifica della Text Encoding Initiative ha accolto in parte la nozione di sottoprospective ed ha sviluppato una serie di costrutti sintattici e semantici in grado di rappresentare adeguatamente fenomeni di sovrapposizione e di parallelismo tra elementi testuali in XML<sup>4</sup>.

---

<sup>4</sup> E tuttavia si deve rilevare che il ricorso ad artifici semantici esula dalla capacità sintattica del linguaggio (e dunque dalla possibilità di verificare la consistenza di un testo codificato con un parser). Nella TEI ad esempio sono stati sviluppati una serie di strumenti di codifica che si basano sui cosiddetti "elementi vuoti" e su "elementi puntatori" per rappresentare caratteristiche testuali, come gli elementi per la dichiarazione e per la rappresentazione di un cosiddetto *feature system* (sistema di caratteristiche) e gli elementi per la rappresentazione di strutture ipertestuali.

### Oltre i limiti della teoria gerarchica

L'ultima versione della teoria gerarchica proposta da Renear si basa su una concezione pluralista e aperta della natura del testo, ben lontana dal platonismo della prima versione, dove il testo era visto come un oggetto unitario, strutturato secondo una singola gerarchia di oggetti di contenuto.

Tuttavia, malgrado questa «apertura» pluralista, il mantenimento della dominante gerarchica non la rende esente da controesempi. È infatti possibile individuare alcuni oggetti testuali (o meglio, oggetti della descrizione testuale) le cui occorrenze si «autosovrappongono», pur appartenendo a una medesima sottoprospettiva analitica: un esempio, decisamente comune nella pratica scientifica in ambito umanistico, è la variante testuale.

La permanenza di controesempi, insieme ai dubbi teorici sollevati da nozioni come quella di prospettiva analitica, sembrano dunque richiedere un ulteriore ripensamento delle teorie della codifica testuale.

Una interessante risposta alle difficoltà determinate dalla teoria gerarchica del testo, è rappresentata dalle formulazioni di natura relativista e costruttivista recentemente emerse nei lavori di alcuni ricercatori. Queste teorie mutuano alcune assunzioni dell'epistemologia relativista postmoderna (in sintonia con la temperie culturale attualmente dominante), avvicinandosi peraltro alle analisi sugli effetti culturali della digitalizzazione del testo formulate da autorevoli studiosi come George Landow [1993 e Landow, Delany, 1993].

Esponenti principali di questa tendenza sono alcuni ricercatori dell'Università di Bergen, che stanno curando il progetto di edizione elettronica dei manoscritti del *Nachlass* di Wittgenstein [Huitfeldt, 1995; Pichler, 1995a e 1995b]. Pur tenendo conto della varietà di posizioni individuali, possiamo dire che questi autori sostengono la necessità di abbandonare l'assunto ontologico che il testo sia un oggetto reale del mondo, dotato di una struttura intrinseca:

Machine-readable texts make it ... clear to us what texts are and what text editing means: Texts are not objectively existing entities which just need to be discovered and presented, but entities which have to be constructed. [Pichler 1995b: 774]

Il testo è dunque una entità che viene costruita e non scoperta e analizzata dalla attività scientifica. In questo quadro i problemi relativi

alla forma strutturale del testo perdono di valore ontologico, e conseguentemente non hanno più senso le obiezioni teoriche e i problemi pratici evocati dalle teorie gerarchiche. Infatti ogni singola prospettiva analitica, anzi più radicalmente, ogni singolo interprete produce una istanza del testo in base alle sue motivazioni e aspettative teoriche e disciplinari. Questa rappresentazione può vantare come unici criteri di adeguatezza la rispondenza effettiva agli interessi di ricerca che l'hanno motivata: ogni oggetto che non rientra nella prospettiva scelta semplicemente non esiste in quel testo, e può essere trascurato.

La teoria costruttivista della codifica testuale offre una soluzione interessante e intellettualmente stimolante alle aporie determinate dalla teoria gerarchica del testo, specialmente nella tematizzazione del ruolo dell'osservatore nei processi di rappresentazione.

D'altra parte, anche le teorie gerarchiche della codifica testuale mostrano un progressivo spostamento dalle prime asserzioni ontologiche realiste sulla natura del testo, verso una visione in cui il punto di vista teorico-disciplinare assume un ruolo costitutivo del testo stesso, o almeno del modo in cui le scienze che si occupano del testo concepiscono la natura del loro oggetto.

Tuttavia, la conclusione secondo cui le aporie della teoria gerarchica e la natura interpretativa della codifica mostrano inequivocabilmente come le rappresentazioni del testo (di cui la codifica elettronica è un esempio speciale), siano il testo stesso

- presenta alcune inconsistenze logiche, ben evidenziate da Allen Renear, al cui scritto rimandiamo<sup>5</sup>;
- rischia di minare i vantaggi metodologici delle applicazioni computazionali nella ricerca umanistica.

Infatti se ogni rappresentazione testuale «costruisce» il suo testo, essa esaurisce il suo compito in se stessa, e soprattutto non è tenuta a osservare criteri di validità e verificabilità intersoggettiva.

---

<sup>5</sup> Si tratta di *Theory and Metatheory in the Development of Text Encoding*, un fondamentale lavoro a cui dobbiamo molto, pubblicato come stimolo di discussione in forma elettronica nella conferenza elettronica *The Monist's. Philosophy and Electronic Publishing*.

Una soluzione teorica alla dicotomia realismo essenzialista/costruttivismo assoluto può emergere dalla nozione di «realismo interno» che Hilary Putnam ha formulato in contesto epistemologico, partendo da una duplice critica tanto al riduzionismo quanto al relativismo:

Noi possiamo e dobbiamo ribadire che alcuni fatti sono là per essere scoperti e non costruiti da noi. Ma questo può essere detto solo dopo che è stato adottato un modo di parlare, un linguaggio, uno «schema concettuale». Parlare di «fatti» senza specificare il linguaggio che deve essere usato è parlare di niente... [Putnam, 1993:149]

In questa visione viene mantenuta la stretta relazione tra linguaggio di rappresentazione e realtà: teorizzare sul (o su aspetti del) mondo senza avere specificato prima un linguaggio non è affatto possibile. Nondimeno il mondo contiene oggetti e fenomeni che esistono indipendentemente dal fatto che qualcuno li descriva. Dipende dalla assunzioni metodologiche che vengono fatte come questi oggetti e fenomeni possano essere descritti. Queste assunzioni a loro volta non sono assunzioni individuali: esse devono essere condivise da una comunità di osservatori che hanno in comune metodologie e pratiche disciplinari, anche se possono divergere sulla interpretazione funzionale di particolari fenomeni. In questo contesto il criterio di identità per sostituzione diventa un criterio di accettabilità razionale non esclusivo: un cambiamento di linguaggio può infatti produrre una diversa visione del mondo, che resta tuttavia altrettanto vera.

In questo senso, ritornando al tema della nostra discussione, non possiamo dire apriori che uno schema di codifica testuale coglie l'essenza del testo più e meglio di un altro in base a un qualche assunto metafisico. Ma neppure si può affermare che ogni rappresentazione è vera in quanto costituisce il suo oggetto testo secondo esigenze specifiche e locali.

La codifica informatica del testo è un linguaggio teorico che permette allo studioso di costruire modelli formali del testo. Ogni modello descrive le caratteristiche del testo a un determinato livello, in base al punto di vista dell'osservatore, ma non coincide con esse. La relazione tra strutture del testo e struttura del modello è complessa e dinamica. Il modello informatico infatti è fondamentalmente un dispositivo euristico: esso consente di fare esperimenti e di verificare processualmente se la descrizione tentata è valida.

Una rappresentazione codificata di un testo dunque 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 metodologica, ed eventualmente di rendere conto dei rapporti tra strutture e fenomeni emergenti da rappresentazioni (punti di vista) diverse.



## La codifica come linguaggio teorico

La diffusione dei sistemi automatici di trattamento e trasferimento dell'informazione interessa ormai ogni ambito dell'agire e del sapere umano. Dall'organizzazione dei processi produttivi, fino all'attività ricreativa, passando per l'educazione e la ricerca scientifica, le macchine informatiche sono ormai una presenza ineliminabile.

Non poteva dunque rimanere immune da tale pervasiva influenza il sapere umanistico e letterario. Nonostante, la resistenza e la diffidenza di moltissimi studiosi, siano, come prevedibile, assai forti, le metodologie e le applicazioni computazionali sono sempre più diffuse, sia nella ricerca sia nella didattica. Il computer d'altronde, che pure nasce per fronteggiare esigenze affatto distanti da quelle delle scienze umane, si è dimostrato un potente strumento per redigere, conservare, diffondere, analizzare i testi.

Tuttavia nel rapporto tra sapere umanistico e informatica, specialmente negli studi letterari, prevale una concezione meramente strumentale, che in parte viene ereditata dalle prime sperimentazioni nel settore: la macchina elettronica (nella quale viene identificato il mondo dell'informatica), in virtù delle sue prestazioni velocistiche, adempie a compiti che precedentemente richiedevano un lungo e meticoloso lavoro, spesso noioso e meccanico. Insomma, un utensile dalle notevoli capacità.

A questa visione strumentale del rapporto tra informatica e ricerca letteraria, si va oggi opponendo una impostazione più radicale, che vede in questo rapporto un vero e proprio salto di paradigma sia dal punto di vista teorico e metodologico sia da quello pratico.

Se per questo secondo aspetto, che riguarda l'organizzazione del lavoro scientifico e didattico, la natura profondamente innovativa delle nuove tecnologie è difficilmente contestabile (basti citare l'uso della rete Internet per la comunicazione scientifica e per la didattica a distanza), i punti di vista sul lato teorico e metodologico sono assai più controversi.

Infatti la visione strumentale a cui abbiamo fatto cenno è ancora profondamente radicata, e si manifesta anche nel lavoro di studiosi che per altro fanno uso di metodologie e strumentazione informatica

per condurre le loro ricerche. Per parte nostra, invece, siamo convinti che l'introduzione di metodologie informatiche determini un vero e proprio cambiamento epistemologico e metodologico nella ricerca letteraria.

Nella nostra cultura la quasi totalità dei testi è veicolata da documenti materiali di varia natura e forma. Per rendere disponibile questo patrimonio attraverso i sistemi elettronici di gestione dell'informazione è necessario dunque effettuare una trasposizione dei testi dal loro supporto originario al nuovo supporto elettronico. Questa memorizzazione informatica del testo, come abbiamo visto, ben lungi dall'essere una mera riproduzione meccanica, presenta invece una natura complessa in cui sono implicati molteplici problemi teorici e pratici, occultati da un approccio tradizionale ai testi e da una concezione idealista del fatto letterario.

### **Lo status epistemologico della codifica testuale**

I linguaggi di markup dichiarativi permettono di predicare l'appartenenza di un dato segmento testuale a una classe di strutture testuali definita dall'utente; potremmo tradurre in linguaggio naturale tali predicazioni con una serie di enunciati del tipo

- questo frammento testuale ha la proprietà "x"

dove "x" sta per una data caratteristica testuale. In tal modo è possibile descrivere formalmente le caratteristiche di un testo in modo indipendente da particolari finalità di trattamento, e da contingenti forme di presentazione grafica su un qualsivoglia supporto fisico (a differenza di altri e più diffusi sistemi di elaborazione testi in uso). Accanto alla plasticità rappresentazionale alcuni di questi linguaggi possono provvedere un insieme di meta-regole (espresse in meta-enunciati) che specificano le relazioni di dipendenza e successione tra gli elementi testuali

I linguaggi di markup dichiarativi, e in particolare SGML e XML, si sono rivelati dei veri e propri strumenti di supporto all'analisi computazionale dei testi. Infatti, la loro plasticità e astrazione non pone limiti alla natura e tipologia delle caratteristiche testuali che si possono codificare. Piuttosto che la struttura editoriale di testo si può rappresentare la sua struttura grammaticale, o la sua struttura retorica.

Non solo: la sintassi del linguaggio di codifica può essere usata per rappresentare le relazioni tra gli elementi strutturali di un testo, a un determinato livello di descrizione.

Le caratteristiche di questi formalismi, insieme alla progressiva diffusione di metodologie computazionali nella ricerca letteraria, ha indotto molti studiosi ad affrontare i fondamenti teorici e metodologici di queste pratiche di ricerca e le loro conseguenze nelle relative pratiche disciplinari.

Alcuni studiosi, come Ian Lancashire e tutti coloro che optano per una semplice riproduzione della sequenza di caratteri del testo mediante coded character set, esprimono forti dubbi circa l'applicazione dei sistemi di codifica dichiarativa in ambito letterario. Essi ritengono che l'astrattezza degli asserti di codifica di un linguaggio di markup dichiarativo introduca una eccessiva dose di soggettività in un processo che si vorrebbe quanto più possibile oggettivo e riproduttivo, orientato dalla riduzione al minimo della perdita di informazioni.

Al di là delle argomentazioni tecniche e pragmatiche particolari (in taluni casi condivisibili, in altri casi frutto di fraintendimenti) che costoro avanzano per sostenere la loro tesi, ciò che caratterizza profondamente tale visione, che potremmo denominare "visione filologica radicale", è l'idea che sia possibile una trascrizione di un testo sul supporto informatico a prescindere da qualsiasi assunto teorico. Questa idea, che fa affidamento su una ipotizzata trasparenza della rappresentazione, permette di conservare fino in fondo la visione strumentale delle applicazioni computazionali in ambito letterario.

Ma si può veramente sostenere che la codifica sia un semplice processo di trascrizione? E, anche se la codifica fosse limitata alla semplice trascrizione, è possibile sostenere che questa sia un processo meccanico, completamente privo di fattori teorico-interpretativi? È possibile in altri termini prescindere dal punto di vista di colui che attua la codifica nel determinare la forma della rappresentazione stessa?

Crediamo che la risposta a tutti questi interrogativi sia negativa. Il trasferimento del testo su supporto informatico propone allo studioso una serie di quesiti teorici (oltre a numerosi problemi pratici) a partire dal momento della decisione su quale particolare oggetto del mondo sia da considerare come "fonte" della memorizzazione.

Consideriamo in primo luogo che cosa precisamente uno studioso possa voler rappresentare sul computer quando si appresta a effettuare la memorizzazione elettronica di un testo. Se analizziamo il termine "testo" del linguaggio ordinario ci accorgiamo che esso si riferisce in

realtà a un insieme di entità diverse, che presentano a loro volta molteplici proprietà non coincidenti:

- a) un documento materiale composto da fogli di carta rilegati (o spillati o semplicemente raccolti insieme da un nastro di carta) che contengono tracce di inchiostro variamente disposte (oltre a eventuali tracce di altri materiali)
- b) un discorso linguistico fissato tramite la scrittura su un documento materiale
- c) un'opera dell'ingegno che viene costituita da quel discorso

Se passiamo poi ad analizzare l'uso del termine testo nell'ambito delle discipline specialistiche che dei testi si occupano, possiamo rintracciare ulteriori significati più o meno simili, come rileva giustamente Sperberg-McQueen nella importante discussione dedicata a questo sul numero interattivo della rivista *Monist*:

- d) lo stato linguistico di un singolo testimone materiale di un'opera (es. il manoscritto *x* del Mattia Pascal; la copia *x* della prima edizione a stampa)
- e) lo stato linguistico di un medesimo testimone di un'opera che presenta diverse lezioni (di cui solo una accolta dall'autore) identificabili
- f) una versione edita di un'opera (ad esempio la prima versione del Mattia Pascal opposta all'ultima revisione). In taluni casi la differenziazione tra le versioni può esser tale da far parlare di testi diversi e dunque di opere diverse (es. la *Gerusalemme liberata* e la *Gerusalemme conquistata*, o *Fermo e Lucia* e *I Promessi Sposi*. Un caso più complesso è costituito dalle due versioni dei Promessi sposi)
- g) una sequenza coerente di enunciati in una lingua naturale

Sebbene questo elenco non sia esaustivo, possiamo rilevare come ciascuna di queste definizioni possa essere applicata a un determinato oggetto empirico che uno studioso ha di fronte nel momento in cui si appresta a farne una rappresentazione elettronica (ovviamente si danno casi in cui non tutte possono essere applicabili). È possibile individuare qualche cosa che accomuna tutte queste diverse concezioni del testo, una sorta di riduzione che possiamo indicare come principio di individuazione di un testo?

Naturalmente il senso (a) sembra non riducibile (si parla di un oggetto materiale) ma possiamo concedere che esso non sia essenziale del concetto di testo, e che la sua comparsa nell'elenco sia un il prodotto di un fraintendimento del linguaggio comune (una sorta di metonimia concettuale). Possiamo dire che l'oggetto materiale su cui sono disposte spazialmente le tracce di inchiostro è quanto la realtà empirica offre ai nostri occhi. Ma solo le tracce in quanto realizzazione grafica di una sequenza di caratteri rivestono per noi interesse: è questa sequenza che da luogo a ciò che chiamiamo testo come scrive Segre:

Se consideriamo i segni grafici (lettere, interpunzione, ecc.) come significanti di suoni, pause, ecc., e riflettiamo sul fatto che questi segni possono essere trascritti più volte e in vari modi (per esempio con grafia e caratteri diversi), restandone immutato il valore, possiamo concludere che il testo è l'invariante, la successione di valori, rispetto alle variabili dei caratteri, della scrittura, ecc. Possiamo anche parlare di significati, se si precisa che si allude a significati grafici, quelli della serie di lettere e segni d'interpunzione che costituiscono il testo. Il testo è dunque una successione fissa di significati grafici. [Segre, 1985: 29]

Il testo, dunque è ciò che permane, *l'invariante*, in ogni operazione di riproduzione materiale della sequenza di simboli grafici. Esso una volta codificato diventa un *tipo (type)* che non ha esistenza materiale in sé, una sequenza virtuale, di volta in volta realizzata da diverse edizioni, copie, trascrizioni, riproduzioni più o meno fedeli, che sono le sue *occorrenze (tokens)*.

Questa definizione di testo come un oggetto astratto allografico sembra fornire un criterio di individuazione di un testo in base al principio di identità per sostituzione. La permanenza del testo come invariante rappresenta la garanzia della sua identificazione e della sua riproducibilità come ha rilevato il filosofo Nelson Goodman:

La differenza nello stile e nelle dimensioni della scrittura e del corpo tipografico, nel colore dell'inchiostro, nel tipo di carta, nel numero e nella composizione delle pagine, nello stato di conservazione ecc. non contano. Conta soltanto quella che potremmo chiamare identità di computazione: la corrispondenza esatta quanto a sequenze di lettere, spazi e segni di punteggiatura. [Goodman, 1968 (1976: 101-102)]

Per la precisione la possibilità di stabilire l'identità di computazione di cui parla Goodman è determinata dalle caratteristiche formali dei

simboli che vengono utilizzati per produrre e riprodurre il testo, i caratteri della scrittura alfabetica:

Per identificare un esemplare dell'opera o per produrre un nuovo esemplare basta verificarne la compitazione o compitare correttamente. Il fatto che un'opera letteraria è composta in una notazione definita, che consiste di certi segni o caratteri che debbono essere combinati in successione, fornisce in effetti il mezzo per distinguere le proprietà costitutive dell'opera da tutte quelle contingenti – cioè a dire, per fissare i tratti necessari e, per ciascuno di essi, i limiti di variazione ammissibili. [Goodman, 1968 (1976: 102)]

Si potrebbe notare che a questa definizione sfugge qualche cosa. Ad esempio, un titolo di paragrafo è certamente una sequenza di caratteri alfabetici. Ma non si può negare che sia una sequenza particolare di caratteri alfabetici: neppure un semplice lettore ingenuo, per dirla con Eco, sarebbe disposto ad accettare che l'essere titolo di una certa sequenza di caratteri non è importante; al contrario molto spesso il titolo svolge una funzione fondamentale per la comprensione di un testo, sia poetico o narrativo sia saggistico o scientifico. Tuttavia si può abbastanza facilmente estendere la definizione di testo per accogliere queste caratteristiche. Ad esempio la teoria della “ordered hierarchy of content objects (OHCO)” [DeRose et al., 1990], è un esempio ottimale di questa estensione. E non a caso tra gli argomenti a sostegno di questa teoria viene citato appunto il criterio di identità per sostituzione.

Ora, anche se ammettiamo che il concetto di testo come invariante allografica o come OHCO forniscano degli argomenti importanti e utili in certi casi per individuare le caratteristiche importanti di un testo, dobbiamo rilevare che essi sono applicabili solo *dopo* una scelta tra le definizioni di testo che nella nostra elencazione vanno da (b) a (g).

Ad esempio se abbiamo stipulato che il testo è la sequenza di compitazione che appare in un dato testimone, possiamo applicare il criterio di Goodman per dire se una data trascrizione è quella corretta. Lo stesso potremmo fare se stipuliamo che per testo intendiamo invece di una edizione critica. Ma non possiamo di norma usare la compitazione di testo come (b) per verificare una trascrizione di un testo come (c). Dobbiamo prima metterci d'accordo circa cosa intendiamo con testo. Solo dopo avere stipulato questo accordo le nostre trascrizioni possono essere confrontate.

Possiamo inoltre notare che almeno nel caso del testo-come-opera potremmo essere propensi a concedere alcune deroghe alla identità di compitazione come criterio di identificazione, deroghe che possono

essere assai larghe (ad esempio, un traduzione dell'opera X è un'opera diversa?). Dunque il processo di trascrizione di un testo sin dall'inizio è orientato da criteri che vengono individuati dal colui che lo effettua.

Un secondo argomento contro la presunta oggettività della rappresentazione riguarda il riconoscimento stesso dei caratteri, su cui si basa il criterio di Goodman. Infatti anche il riconoscimento di un carattere su un testo scritto, implica assunzioni teoriche e interpretazione: l'assunzione che una data traccia grafica "A" sia un token di un data classe astratta di tracce che identifichiamo come il carattere "a". Il riconoscimento dei caratteri di un testo dunque è una abduzione ipercodificata, nella terminologia di Eco. Se nella maggioranza dei casi chiunque è in grado di effettuare in modo automatico questa abduzione, si danno molti casi in cui questo non è facile, o è del tutto impossibile, e vengono richiesti ulteriori sforzi interpretativi: "riconoscere un fenomeno dato come il token di un dato type presuppone alcune ipotesi sul contesto espressivo e sul co-testo discorsivo" [Eco, 1990: 237]. Non solo. Come rileva Daniel Dennet, se abbiamo una iscrizione di questo tipo:

TAE CAT

siamo in grado di riconoscerla solo se conosciamo l'inglese (lingua e ortografia) e se assumiamo che il suo autore volesse dire qualcosa di sensato.

We rely on the norms for the formation of letters, on the one hand, and on the other we have expectations about the likelihood that this inscription was produced with some communicative *intent*. We try to see a message in any string of letters, and in this instance we succeed. Of course *maybe* the author of this inscription *meant* to write "the cht" (or "tae cat")--in which case we have erred in our headlong interpretation. [Dennet, 1990].

E come rileva lo stesso Dennet, possiamo benissimo sbagliarci nell'interpretare. Casi come questi sono tipici nella pratica filologica (si noti che in molti casi nemmeno allargando il contesto si potrebbe arrivare a una disambiguazione certa).

In taluni casi il riconoscimento dei caratteri su un supporto scritto comporta la costruzione vera e propria di nuovi codici (ad esempio se

si stanno decifrando delle iscrizioni sconosciute). Questa consapevolezza è ben presente nel lavoro di filologi e codicologi. Essi probabilmente non accetterebbero in linea di principio l'idea che è sempre possibile predicare in modo non ambiguo l'appartenenza di una certa "traccia" su un supporto testuale fisico a una certa classe di iscrizioni che chiamiamo "carattere":

Even the most diplomatic of editions of a manuscript source [...] enshrines in the authority of print one set of editorial assumptions about emphasis, illegibility, and even the interpretation of the letters of the text itself [...]

What is illegible to one editor may be perfectly plain to another [...]; what one person regards as a standard abbreviation may require explicit expansion for another. Experts may agree only that the decision as to whether a "u" or a "v" is intended at this point is not decidable. [Burnard, Greenstein, 1995:137]

Insomma per riconoscere un carattere abbiamo dovuto ricorrere a diversi livelli di competenza:

- la conoscenza del codice notazionale della scrittura;
- la conoscenza del codice verbale;
- l'intervento di competenze contestuali e circostanziali.
- l'attribuzione di intenzioni comunicative al testo o all'autore

Un discorso simile potrebbe essere effettuato per quella classe di oggetti previsti dalla teoria OHCO. Essa prevede infatti la presenza di una struttura gerarchica e ordinata di oggetti chiaramente riconoscibili perché connesse in modo causale con alcune caratteristiche fisiche (tipografiche e compositive) del documento materiale. Tuttavia tale identificazione è spesso assai controversa. Ad esempio in alcuni testi sono presenti degli spazi bianchi che separano sequenze di paragrafi, ma che non sembrano avere le caratteristiche delle normali suddivisioni di capitolo, sottocapitolo etc. La scelta di rappresentare queste separazioni come identificatori di sezioni strutturali piuttosto che come occorrenze localizzate di spaziature può essere presa a ragion veduta dallo studioso che sta attuando la codifica solo dopo avere analizzato e interpretato la struttura complessiva del testo e il suo contenuto linguistico.

## La codifica come linguaggio teorico

Le riflessioni che abbiamo condotto finora, sebbene non possano dirsi sistematiche né tantomeno esaustive, ci hanno mostrato come, anche se volessimo limitare la rappresentazione di un testo alle caratteristiche che hanno una controparte osservazionale (come ad esempio la rappresentazione dei caratteri che costituiscono il testo) ci troveremo di fronte a scelte teoriche e non potremmo fare a meno di introdurre elementi interpretativi.

Insomma, come hanno rilevato Burnard e Greenstein, non esiste nessun fondamento per distinguere in modo assoluto o essenziale una codifica (buona) puramente riproduttiva da una codifica (cattiva) interpretazionale:

... we are unconvinced that there can ever be a fundamental distinction between [transcription and interpretation], other than established consensus. A neutral transcription is, we suggest, merely one in which the set of interpretive distinctions made happens to coincide with the set of such distinctions most people would wish to make most of the time; an analytic or interpretative transcription is one in which the set of distinctions made is peculiar to some specific goal or agenda. It is hard to establish a more exact distinction, even with the aid of a Martian observer. [Greenstein - Burnard, 1995:146]

L'individuazione delle varie tipologie di informazioni veicolate nel testo, delle modalità di rappresentazione informatica delle stesse e dei limiti da porre a tale operazione, per non cadere in regressi *ad infinitum*, è sempre il risultato di una serie di scelte teoriche e di atti interpretativi, che vengono incorporati esplicitamente o implicitamente nella rappresentazione stessa. La codifica elettronica di un testo, in quanto rappresentazione di un testo e delle sue caratteristiche mediante un linguaggio formale, si colloca interamente all'interno del processo analitico-interpretativo, né prima né dopo.

L'attività di codifica, dunque, assume una funzione metodologica nell'ambito delle discipline che si occupano del testo, e il linguaggio di codifica adottato può essere considerato come un *linguaggio teorico*. Tale linguaggio, al pari di ogni altro linguaggio teorico, è utilizzato dallo studioso per costruire teorie o modelli dei fenomeni testuali di suo interesse e per esplicitare le sue ipotesi interpretative su un certo oggetto di studio (il testo e le sue caratteristiche a un determinato livello di analisi).

Per questa ragione l'adozione di sistemi di rappresentazione informatica dei testi a mio parere costituisce un notevole cambiamento metodologico nelle scienze testuali.

Il linguaggio teorico delle scienze testuali è sempre stato basato sul linguaggio ordinario, usato in funzione metalinguistica. Le oscillazioni semantiche e sintattiche del linguaggio dunque, in un certo senso mascherano l'incerto status teorico degli oggetti studiati nel corso dell'analisi. L'adozione di un linguaggio teorico formalizzato impone allo studioso la definizione formale degli oggetti di studio e la formulazione delle inferenze e delle interpretazioni, in modo rigoroso esplicito e intersoggettivamente verificabile, mediante i costrutti del linguaggio stesso.

In questo contesto il problema della codifica non è più quello di riuscire a trovare un sistema per rappresentare il testo che sia oggettivo in senso metafisico, ma piuttosto quello di individuare o sviluppare un sistema di codifica abbastanza potente da permettere a ogni studioso, da qualsiasi punto di vista disciplinare, di rappresentare le caratteristiche testuali che lo interessano e di poter esplicitare le sue interpretazioni sul ruolo di tali caratteristiche. Non solo: come ricordavamo all'inizio la rappresentazione informatica di un testo è il presupposto necessario di qualche procedura di trattamento. Occorre dunque tenere presente nella rappresentazione del testo anche i possibili processi applicativi a cui esso può essere sottoposto. In altri termini, deve essere possibile poter specificare una simulazione informatica di un processo di trattamento, che operi sulla rappresentazione del testo.

Si aprono qui una serie di considerazioni e problemi relativi ai formalismi che attualmente sono stati sviluppati per effettuare la codifica. Non posso per ovvi motivi di tempo soffermarmi su tali temi, anche perché studiosi dei ben maggiore esperienza hanno fornito importanti contributi teorici e pratici al riguardo. Mi limito pertanto a citare e sottoscrivere quanto scritto dal professor Sperberg McQueen in intervento al Forum *Philosophy and electronic publishing* organizzato dalla rivista *Monist*.

My reason for seeing SGML as a great step forward was, as I recall it, that it allowed a whole new repertoire of operations on the document conceived of as a tree, without losing the ability to operate on the document conceived of as a linear data stream. So yes, I certainly did have a particular set of operations in mind: two sets, to be precise. And any scheme which supports both sets of operations as well as SGML or any other OHCO-plus-hypertext scheme will be a text encoding scheme to take seriously.

Desidero piuttosto dedicare le ultime riflessioni alcune conseguenze di ordine “ontologico” implicate nella teoria della codifica, stimolato dalle fondamentali osservazioni che Allen Renar ha dedicato a questo tema.

Alcuni studiosi, partendo da considerazioni in parte simili a quelle che ho espresso in questa comunicazione, hanno assunto delle posizioni espressamente costruttiviste riguardo la natura ontologica dei testi. In particolare Alois Pichler del Wittgenstein Archivium è stato assai esplicito:

Machine-readable texts make it ... clear to us what texts are and what text editing means: Texts are not objectively existing entities which just need to be discovered and presented, but entities which have to be constructed. [Pichler 1995b:774]

Questa tesi, che Pichler ribadisce anche nel suo intervento al forum di *Monist* ricordato poco fa, può facilmente sfociare in una forma di relativismo assoluto che ha conseguenze assai poco desiderabili. Se ogni rappresentazione testuale “costruisce” il suo testo, essa esaurisce il suo compito in se stessa, e soprattutto non è tenuta a osservare criteri di validità e verificabilità intersoggettiva.

Per quanto mi riguarda, non credo che adottare una epistemologia interpretazionale nei fatti della codifica testuale implichi necessariamente l'accoglimento di qualche forma di relativismo assoluto né di un antirealismo metafisico rispetto ai testi. Né d'altra parte riterrei accettabile una metafisica realista che presupponga una sorta di visione dell'Occhio di Dio (ben più potente di quello dei marziano che evocavano Burnard e Greenstein)

Mi sembra che il filosofo Hilary Putnam fornisca una interessante alternativa con la sua idea di “realismo interno”:

Noi possiamo e dobbiamo ribadire che alcuni fatti sono là per essere scoperti e non costruiti da noi. Ma questo può essere detto solo dopo che è stato adottato un modo di parlare, un linguaggio, uno “schema concettuale”. Parlare di “fatti” senza specificare il linguaggio che deve essere usato è parlare di niente: l'uso della parola fatto non è fissato dal mondo stesso, più di quanto non lo sia l'uso delle parole “esiste” e “oggetto” [Putnam, 1993: 149]

Il mondo contiene oggetti e fenomeni (tra cui i testi) che esistono indipendentemente dal fatto che qualcuno li descriva. Ma per dire che cosa e come siano questi oggetti dobbiamo metterci d'accordo su come usare i concetti e i termini del nostro linguaggio. Mettersi

d'accordo significa molte cose come descrivere in che modo un linguaggio viene utilizzato da una comunità di osservatori/parlanti, in che modo si ritiene accettabile usare i concetti, quali metodologie e criteri sono utilizzati per asserire la verità degli enunciati, e così via.

In questo senso, non possiamo dire *apriori* se uno schema di codifica testuale coglie l'essenza del testo più e meglio di un altro in base a un qualche principio metafisico. Ma neppure si può affermare che ogni rappresentazione è vera in quanto costituisce il suo oggetto testo secondo esigenze specifiche e locali.

La rappresentazione di un dato testo dipende dalle assunzioni teoriche e dalle interpretazioni che vengono fatte da uno studioso. Queste assunzioni a loro volta non sono individuali: esse sono condivise da una comunità di studiosi che hanno in comune metodologie e pratiche disciplinari, ontologie pratiche, criteri di accettabilità razionale, anche se possono divergere sulla interpretazione di particolari fenomeni.

Una rappresentazione codificata di un testo è "corretta" se è internamente coerente, se è accettabile razionalmente nell'ambito di una teoria, e se permette di rappresentare ed spiegare i fenomeni testuali rilevanti nel contesto di quella teoria o prospettiva metodologica. Potrebbe sembrare forse troppo poco, ma è quanto siamo in grado di fare.

## La rappresentazione complessa di testi complessi

La codifica elettronica dei testi ha rappresentato uno dei temi fondamentali della riflessione e della sperimentazione nel dominio proteico e multiforme dell'Informatica umanistica.

D'altra parte se solo si riflette un poco sul tema è facile trovare giustificazione per questa centralità: i testi, nelle loro varie forme materiali, sono senza dubbio sia oggetto principale sia strumento di studio di buona parte delle discipline umanistiche. Dunque l'adozione di metodologie informatiche nello studio dei testi richiede in primo luogo di provvedere a una loro adeguata rappresentazione in formato digitale. Ora è ben noto come questa apparente ovvietà abbia rivelato un intreccio di complesse questioni teoriche e pratiche, appunto oggetto della vasta bibliografia cui si faceva cenno poco fa. Il nodo della questione deriva dalla tensione che si stabilisce tra i due poli della rappresentazione: testo e rappresentazione digitale. Per entrambi infatti si può predicare l'attributo della complessità sebbene con accezioni non perfettamente sovrapponibili.

Il testo è un oggetto complesso in quanto è in grado di veicolare significato (e di rivestire dunque interesse scientifico) su più livelli strutturali, e cosa ancora più rilevante attraverso l'instaurazione di molteplici relazioni tra più livelli. Come rileva Jerome McGann questa complessità viene esplicitamente tematizzata dai testi letterari e poetici, i quali, dunque, assumono un ruolo privilegiato nella nostra comprensione della comunicazione testuale:

Considered strictly in term of bibliographical codes, then, poetical works epitomize a crucial expressive feature of textuality in general: that it can be seen to organize itself in terms of various relational segmentations and meta-segmentations. Some elementary segmentations are sentences, paragraphs, chapters; in verse lines, inter- and intralinear forms (rhyme, for example, and metrical forms) stanzas, cantos; in the page, the opening, the book. These segmentations may be usefully traced to the level of the individual character and, in general, to font and typeface design. [McGann, 2001: 183]

Non possediamo nessuna teoria sufficientemente completa del testo, e nemmeno teorie parziali di uno o più livelli che si possano defi-

nire in senso stretto formali (con le dovute eccezioni: alcune teorie linguistiche ad esempio, o alcune teorie meccaniciste della critica testuale). Il fatto è che la rappresentazione (e a maggior ragione l'elaborazione) informatica è ontologicamente formale in senso stretto. La storia della codifica informatica dei testi consiste nel tentativo di superare questo duplice divario concettuale:

- 1) sviluppare teorie e modelli formali del testo (o di alcuni suoi livelli descrittivi)
- 2) individuare formalismi atti a esprimerli in modo computazionalmente accettabile<sup>1</sup>

Come è noto la soluzione considerata teoricamente ottimale e ampiamente diffusa nella pratica dalla comunità scientifica consiste nell'adozione dei markup language descrittivi basati su XML (o sul suo diretto predecessore SGML). Lo sviluppo e la diffusione della Text Encoding Initiative ha sancito questa scelta, superando non poche diffidenze, resistenze e persino ostracismi culturali. A distanza di venti anni dall'avvio del progetto oggi si può dire che la TEI rappresenta un punto di riferimento per tutte le imprese volte a digitalizzare risorse testuali in ambito umanistico per fini di ricerca e di conservazione digitale.

La diffusione di XML e della TEI nella pratica concreta della digitalizzazione in area umanistica è stata affiancata da una notevole quantità di lavori volti a fornire giustificazioni teoriche per questa preminenza. Questa attività teorica ha dato luogo a teorie metafisicamente impegnative come la ormai nota *OHCO* [cfr. *supra*]: alla domanda che cosa è un testo veramente rispondiamo che un testo è un oggetto linguistico astratto organizzato secondo una struttura gerarchica ordinata di oggetti di contenuto. La fallacia di questa asserzione generale è stata ben presto riconosciuta dai suoi stessi estensori, che

---

<sup>1</sup> Si noti che per paradosso linguistico un formalismo (o un algoritmo) è tanto più utilizzabile quanto è meno complesso nel senso assunto dal termine nella teoria della complessità computazionale. Dove la complessità è una misura delle risorse temporali o spaziali (considerate indipendentemente dalla potenza e velocità di calcolo della macchina) necessarie a elaborarlo. Detto in termini semplici, è inutile avere formalismi la cui computazione potrebbe essere irragionevolmente lunga o esosa, ovvero intrattabile.

ne hanno proposto negli anni diverse revisioni più aperte e pluraliste, ma l'idea di una preminenza della struttura gerarchica nella testualità ha mantenuto un ruolo descrittivo ed esplicativo essenziale.

La ragione di tanto attaccamento all'idea di struttura gerarchica ovviamente non è immotivata. Il fatto è che XML (e SGML) può essere considerato sia un formalismo sia un modello di dati espresso da quel formalismo, e tale (meta)modello è appunto un *albero ordinato etichettato*. In altri termini XML considerato come linguaggio di modellizzazione può esprimere solo modelli la cui struttura è un albero. Di conseguenza tutte le applicazioni XML standard sono in grado di elaborare (validare, trasformare, formattare) strutture dati ad albero.

Considerato come puro formalismo sintattico XML può essere utilizzato per rappresentare linearmente (serializzare) anche strutture e modelli di dati non strettamente gerarchici, e in questo di fatto risiede gran parte del suo successo come standard informatico. Tuttavia se si adotta XML come puro linguaggio di serializzazione le possibilità di elaborazione fornite da applicazioni XML standard su tali dati sono assai ristrette (di fatto ristrette a ciò che può essere trattato come albero). Il prezzo costituito dall'adozione di un modello di dati così vincolante, d'altra parte, paga il vantaggio di potere validare in modo automatico ogni istanza di dati rispetto al modello mediante algoritmi generali ben conosciuti e computazionalmente trattabili, ciò che a sua volta consente di costruire sistemi di elaborazione degli stessi dati consistenti ed efficaci<sup>2</sup>.

Naturalmente se si considerano le finalità applicative prioritarie per i quali sia SGML sia XML sono stati sviluppati, si può ben comprendere come nella comunità informatica non si sia avuta alcuna remora a considerare questo vantaggio un ben valido contraccambio per il prezzo pagato. In effetti nell'elaborazione informatica dei documenti in ambito commerciale e industriale (cioè volta a produrre, archiviare e disseminare in modo efficiente e sicuro documentazione in formato digitale, non a descrivere testi complessi in modo complesso per fini

---

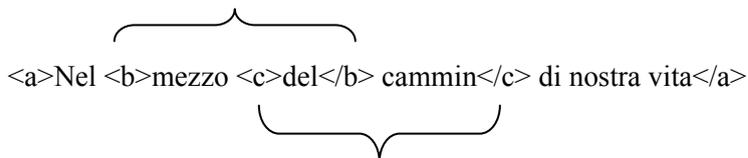
<sup>2</sup> A dire il vero l'affermazione che XML possa modellizzare adeguatamente solo alberi può essere attenuata in virtù della capacità di esprimere relazioni non gerarchiche tra elementi mediante coppie di attributi ID/IDREF. Questi costrutti permettono sia di validare sia di elaborare relazioni strutturali non strettamente gerarchiche tra gli elementi di un documento XML. Tuttavia la capacità espressiva di questi costrutti è ovviamente assai limitata.

complessi), tale prezzo era in realtà nullo: ogni documento testuale può essere ragionevolmente descritto come un albero da un qualche punto di vista, basta adottare il livello giusto di segmentazione e astrazione. Ma anche considerando testi complessi come quelli oggetto della ricerca umanistica è innegabile che il modello gerarchico si presti naturalmente a descrivere la struttura di numerosi livelli testuali scientificamente rilevanti: i livelli editoriale, morfosintattico, metrico, tanto per fare alcuni esempi, sono in linea generale descrivibili come strutture gerarchiche ordinate.

I problemi infatti sono emersi solo quando la comunità degli studiosi di scienze del testo con tutto il loro bagaglio di teorie e interpretazioni del concetto di testo hanno pensato di eleggere SGML e XML a formalismi generali per la rappresentazione complessa dei loro oggetti di analisi. Le manifestazioni di queste difficoltà sono state comunemente rubricate come il problema delle gerarchie sovrapposte (*overlapping hierarchies*, OH d'ora in poi)

### **Il problema delle gerarchie sovrapposte**

La prima enunciazione del problema e dei suoi casi esemplari con le relative “pseudo-soluzioni” risale al noto articolo di Barnard e altri «Hierarchical Encoding of Text: Technical Problems and SGML Solutions» [Barnard et al., 1988]. In termini semplici il problema OH dal punto di vista sintattico consiste nel fatto che, dati due oggetti logici presenti in un testo, le coppie di tag bilanciati che li rappresentano non si annidano propriamente ma si sovrappongono.


  
 <a>Nel <b>mezzo <c>del</b> cammin</c> di nostra vita</a>

Questo corrisponde alla presenza di almeno due sotto-alberi (uno con il ramo  $a \rightarrow b$  e uno con quello  $a \rightarrow c$ ) che descrivono contempora-

neamente la struttura del documento. Tale situazione è sintatticamente e semanticamente vietata in XML<sup>3</sup>.

Dal punto di vista concettuale il problema OH è un sottoinsieme del più generale problema della complessità testuale, come a più riprese hanno fatto notare autori come Dino Bozzetti e Jerome McGann:

My own comparison is itself a kind of joke, of course, for an SGML model of the world of textualities pales in comprehensiveness before the Newtonian model of the physical world. But the outrageousness of the comparison in each case helps to clarify the situation. No autopoietic process or form can be simulated under the horizon of a structural model like SGML, not even topic maps. We see this very clearly when we observe the inability of a derivative model like TEI to render the forms and functions of traditional textual documents [McGann, 2004]

Il modello SGML/XML, infatti, presuppone che esistano almeno due livelli testuali per i quali si possa individuare e formalizzare una struttura chiara di elementi testuali e di relazioni formali tra questi elementi, o detto altrimenti, che esista un modello formale di quei livelli testuali (ed eventualmente delle relazioni tra tali livelli). Ma questo come già si notava è un presupposto irrinunciabile per adottare metodologie informatiche nella rappresentazione e analisi computazionale dei testi. Se esistono proprietà dei testi irriducibili a qualsiasi formalizzazione anche minimale, allora queste non possono per definizione essere rappresentate e trattate con metodi computazionali<sup>4</sup>.

---

<sup>3</sup> L'uso di coppie bilanciate e annidate di simboli è la strategia standard di rappresentazione lineare di una struttura gerarchica: si pensi alla sintassi delle espressioni algebriche. XML adotta questo meccanismo sintattico sottoforma di coppie di tag.

<sup>4</sup> Si noti che questo vale anche per la creazione di facsimili digitali in formato immagine delle pagine di un libro. Infatti:

1. nonostante le apparenze anche l'immagine digitale è un modello formale regolato da precise proprietà matematiche
2. tutto ciò che si può fare su una immagine digitale con metodologie strettamente computazionali deriva dalle proprietà formali del modello e dalla individuazione di algoritmi che possano manipolare tali proprietà

In effetti i limiti attuali dell'elaborazione di immagini permettono di fare molto poco con tali facsimili: tutta l'elaborazione veramente interessante che si può fare su di esse viene eseguita dall'agente umano che vi accede attraverso un qualche dispositivo e supporto di output. Ovvero, l'immagine digitale di un manoscritto è veramente

## Una gerarchia per le gerarchie sovrapposte

In realtà nella categoria delle gerarchie sovrapposte si possono distinguere diversi sottoproblemi di complessità crescente (De Rose 2004). Possiamo insomma distinguere una “gerarchia delle gerarchie sovrapposte”.

1) Il caso più semplice e comune è quello della compresenza di due o più strutture (livelli) gerarchiche i cui elementi si sovrappongono. Nell’ambito di questa classe è possibile individuare una scala crescente di complessità di trattamento in ragione del numero di livelli compresenti e del numero di elementi di un dato livello che vengono attraversati da un elemento di un altro livello

In questo esempio (dall’*Antigone* di Alfieri) la struttura metrica si sovrappone alla struttura editoriale del testo drammatico:

SCENA III  
ARGIA, ANTIGONE

**Argia**

Una infelice io sono.

**Antigone**

In queste soglie  
che fai? che cerchi in sì tard’ora?

**Argia**

Io... cerco...  
... d’Antigone...

**Antigone**

Perché? — Ma tu, chi sei?  
Antigone conosci? a lei se’ nota?  
che hai seco a far? che hai tu comun con essa?

Si noti che volendo codificare contestualmente anche la struttura linguistico-grammaticale si genererebbero ulteriori sovrapposizioni anche multiple.

utilizzabile solo dallo studioso che la osserva e la interpreta, né più né meno che se osservasse l’originale o un facsimile su carta o su microfilm. Ovviamente questo è vero ora stante l’attuale livello di sviluppo delle tecnologie di *image processing*. Sul futuro è meglio non fare previsioni.

2) Un caso concettualmente più complicato è quello di un elemento appartenente a una gerarchia che si estende oltre i confini dell'elemento in cui inizia o persino di uno dei suoi predecessori (citazioni, annotazioni, fenomeni materiali nella trascrizioni di fonti primarie).

In questo esempio (dalla *Commedia*) il discorso diretto si sovrappone con la segmentazione in versi

Rispuosemi: “Non omo, omo già fui,  
e li parenti miei furon lombardi,  
mantoani per patria ambedui.

3) Tecnicamente simile ma concettualmente distinto il caso di elementi composti da segmenti discontinui e non contigui (di nuovo discorsi diretti, annotazioni, cancellature o aggiunte, fenomeni materiali), come nel seguente frammento dal *Turno* di Pirandello dove l'enunciato diretto di donna Rosa è inframmezzato al livello diegetico della narrazione:

– In nome del Padre, del Figliuolo e dello Spirito Santo,  
– si lamentava intanto, in casa, la moglie  
del Ravi, la si-donna Rosa, accennando il segno della  
croce con un gesto che le era abituale e che ripeteva  
ogni qual volta si sentiva infastidita e urtata nella  
gravezza della sua gialla carne inerte: – Lasciatelo  
fare. Ciò che fa Marcantonio, per me, è ben fatto,–  
diceva ai parenti che sottovoce le facevan notare  
la mostruosità di quel progetto di nozze.

4) Infine il caso più complesso che si verifica quando un dato elemento può auto-sovrapporsi illimitatamente.

Nell'ottavo periodo del capitolo VIII del *De Principatibus* troviamo ad esempio le seguenti varianti:

- lasciato parte della sua gente {( <alla [defesa) della> obsidione} }, con le
- lasciato parte della sua gente <alla offesa delle> obsidione, con le
- lasciato parte della sua gente alla [difesa di quella], con le
- lasciato parte della sua gente (difesa) della obsidione, con le
- gente {allobsidione}, con le

Le varie parentesi che abbiamo inserito mostrano i rapporti di variazione tra il lemma e le varie lezioni attestate. Come è evidente un ipotetico elemento XML `<variante>` darebbe luogo a numerose auto-sovrapposizioni.

### **Soluzioni per il problema delle gerarchie sovrapposte**

La soluzione al problema OH, come si diceva prima, potrebbe non esaurire completamente le difficoltà della formalizzazione digitale del testo. Tuttavia esso rappresenta una consistente “minaccia teorica” interna al paradigma della codifica testuale basata sui linguaggi di markup strutturati, la cui rilevanza è ben chiara a uno dei massimi esponenti di questa scuola, Michael Sperberg-McQueen:

It is an interesting problem because it is the biggest problem remaining in the residue. If we have a set of quantitative observations, and we try to fit a line to them, it is good practice to look systematically at the difference between the values predicted by our equation (our theory) and the values actually observed; the set of these differences is the residue .... In the context of SGML and XML, overlap is a residual problem [Sperberg-McQueen, 2002]

Non è un caso che negli ultimi dieci anni, proprio in parallelo con l'inarrestabile diffusione di XML nel mondo dell'elaborazione testuale (e non solo) e della TEI nella comunità umanistica si sono moltiplicati i tentativi di trovare delle soluzioni definitive al problema. Le soluzioni proposte si possono dividere in due classi: soluzioni interne e soluzioni esterne al paradigma XML.

Prima di passare in rassegna queste soluzioni, tuttavia, è opportuno individuare i criteri per valutarne l'appropriatezza e l'accettabilità. Il primo criterio da prendere in considerazione ovviamente è la completezza e congruenza della rappresentazione con i fenomeni strutturali rappresentati (isomorfismo del modello rispetto al sistema reale). Ma come giustamente ha rilevato Steve De Rose, il solo criterio della congruità rappresentazionale è insufficiente. È infatti sempre possibile indebolire i vincoli di un formalismo fino a renderlo abbastanza “accolgente” da permettergli di modellizzare qualsiasi fenomeno testuale, ma in questo modo la sua efficacia euristica tende a degradare fino ad annullarsi.

Per questo alla completezza e congruenza va affiancata una serie di ulteriori criteri valutativi di natura teorica, tecnica e pragmatica, proposti di recente dallo stesso De Rose che qui riassumiamo:

1. Leggibilità e facilità di comprensione da parte di un utente umano (sono esclusi dunque tutti i formati binari)
2. Facilità di manutenzione e modifica
3. Disponibilità di implementazioni software
4. Compatibilità sintattica con XML
5. Facilità di validazione (eventualmente sopravanzando le capacità di validazione di un *parser* XML standard)
6. Possibilità di validazione incrociata tra gerarchie diverse
7. Possibilità di formattazione ed elaborazione grafica e presentazionale
8. Possibilità di estrapolare molteplici viste basate su uno o più tra le gerarchie presenti
9. Possibilità di estrapolare sottoinsiemi gerarchici delle caratteristiche testuali
10. Continuità del contenuto testuale serializzato [De Rose, 2004]

### **Le soluzioni interne al paradigma XML**

Nella classe delle soluzioni interne al paradigma XML rientrano artifici sintattici che mantengono la conformità a XML (e dunque le proprietà di buona formazione e validità dei documenti prodotti) ma che di fatto si basano sulla possibilità di utilizzare il metalinguaggio (o una sua applicazione) come puro formalismo di serializzazione per modelli di dati non gerarchici (fondamentalmente grafi orientati connessi, di cui gli alberi sono un sottoinsieme). Di conseguenza sebbene le seguenti strategie siano tutte sintatticamente conformi a XML non è vero che esse siano sempre semanticamente trattabili con processori XML standard.

#### *Segmentazione*

Un elemento logico che si sovrappone ai confini di un'altro (o di più altri) viene diviso in due (o  $n$ ) elementi XML dello stesso tipo correlati mediante apposti attributi. Ad esempio, in questo modo si po-

trebbe trattare il caso delle strutture editoriali e metriche della tragedia alfieriana vista sopra:

```
<div2 type="scena">
<head>SCENA III</head>
<stage>ARGIA, ANTIGONE</stage>
<sp><speaker><emph>Argia</emph></speaker>
<l part="I">Una infelice io sono.</l>
</sp>
<sp><speaker><emph>Antigone</emph></speaker>
<l part="F">In queste soglie</l>
<l part="I">che fai? che cerchi in sì tard'ora?</l>
</sp>
<sp><speaker><emph>Argia</emph></speaker>
<l part="F">Io... cerco...</l>
<l part="I">... d'Antigone...</l>
</sp>
<sp><speaker><emph>Antigone</emph></speaker>
<l part="F">Perché? – Ma tu, chi sei?</l>
<l>Antigone conosci? a lei se' nota?</l>
<l>che hai seco a far? che hai tu comun con
essa?</l>
</sp>
```

Questa soluzione permetterebbe di risolvere sovrapposizioni di tipo (1), (2) e (3), e può essere parzialmente validata mediante un oculato uso di attributi ID/IDREF. Tuttavia è impossibile da utilizzare se sussistono molteplici sovrapposizioni sulla stessa sequenza di caratteri. Inoltre è alquanto verboso (specialmente se si devono collegare più di due segmenti) e di conseguenza difficilmente leggibile da parte di un utente umano e di complessa manutenzione.

Ma soprattutto costringe a individuare una e una sola gerarchia come primaria e determina una moltiplicazione arbitraria degli elementi sintattici rispetto agli oggetti logici rappresentati

### *Elementi di congiunzione*

Si tratta di una evoluzione del sistema basato sulla segmentazione. Consiste nell'introduzione di un elemento XML (elemento di congiunzione) con la funzione metatestuale di esprimere l'unità logica di un fenomeno testuale rappresentato da più elementi XML distinti (non

necessariamente dello stesso tipo). Ad esempio il caso di un discorso diretto frammentario si potrebbe gestire in questo modo:

```
<p>- <q id="q1">In nome del Padre, del Figliuolo e
dello Spirito Santo,</q> - si lamentava intanto, in
casa, la moglie del Ravì, la si-donna Rosa,
accennando il segno della croce con un gesto che le
era abituale e che ripeteva ogni qual volta si
sentiva infastidita e urtata nella gravezza della
sua gialla carne inerte: - <q id="q2">Lasciatelo
fare. Ciò che fa Marcantonio, per me, è ben
fatto,</q>- diceva ai parenti che sottovoce le
facevan notare la mostruosità di quel progetto di
nozze.</p>
```

...

```
<join targets="q1 q2" result="q"/>
```

Questa tecnica consente di trattare in linea teorica ogni caso di sovrapposizione e di non contiguità, di ordinamento inverso e di relazione n-aria tra oggetti testuali. Tuttavia si presenta anch'esso assai prolisso, specialmente se i livelli di sovrapposizione sono numerosi, e di difficile manutenzione. Esso inoltre presuppone che il testo sia segmentare esplicitamente in elementi base da unire, ciò che pèoytrebbe determinare una moltiplicazione del markup a fini puramente sintattici. Questo ultimo problema può essere risolto parzialmente adottando gli schemi di puntamento basati su *range* di caratteri previsto nello standard XPointer ma in questo caso l'elaborazione dei riferimenti richiederebbe l'uso di software *ad hoc*.

### *Markup esterno (stand-off markup)*

Le tecniche basate su markup esterno sono di fatto identiche a quelle basate su elementi *join*. La differenza consiste nel fatto che in questo caso gli elementi che esprimono il collegamento di segmenti testuali nel documento XML base sono in un documento XML esterno. Riformulando l'esempio precedente:

```
<p>- <q id="q1">In nome del Padre, del Figliuolo e
dello Spirito Santo,</q> - si lamentava intanto, in
casa, la moglie del Ravì, la si-donna Rosa,
accennando il segno della croce con un gesto che le
era abituale e che ripeteva ogni qual volta si
```

sentiva infastidita e urtata nella gravezza della sua gialla carne inerte: - <q id="q2">Lasciatelo fare. Ciò che fa Marcantonio, per me, è ben fatto,</q>- diceva ai parenti che sottovoce le facevan notare la mostruosità di quel progetto di nozze.</p>

In un file esterno avremo i seguenti elementi:

```
<link xlink:type='extended'>
<anchor xlink:type='locator' xlink:role='quote'
href="id(q1)">
<anchor xlink:type='locator' xlink:role='quote'
href="id(q1)">
</link>
```

Il fatto di poter collocare i collegamenti in un documento autonomo consente di adottare un linguaggio XML per descrivere una struttura principale del documento base e uno diverso per la rappresentazione dei livelli di descrizione ulteriori. Ad esempio si può usare la TEI per la codifica del documento e XTM (*XML Topic Maps*) o RDF per esprimere le relazioni tra gli oggetti testuali. Questa tecnica ha il vantaggio di poter disporre di modelli di dati e di sistemi di elaborazione complessi (quali sistemi di deduzione per RDF o Topic Map). Naturalmente il lato opposto della medaglia è l'aumento della complessità di gestione complessiva del sistema. Per il resto si possono portare gli stessi argomenti avanzati sul metodo basata su elementi di congiunzione.

### *Elementi milestone*

Un elemento *milestone* è un elemento XML vuoto che segnala un punto monodimensionale in un documento XML. Questi elementi ovviamente non presentano problemi di sovrapposizione sintattica e si possono collocare liberamente in un documento XML (salvo restrizioni espresse da uno schema o da una DTD). In questo modo è possibile segnalare i confini di qualsiasi struttura documentale.

Questa strategia è ampiamente utilizzata nella TEI per veicolare le indicazioni sulla messa in pagina di un testo nel documento fonte da cui è stato memorizzato ( si pensi a elementi come <pb>, <cb>, <lb>

per indicare rispettivamente il salto di pagina, colonna e riga), come nel seguente esempio (*La notte*, da *I canti orfici* di Dino Campana)

```
<div2 type="poemetto">
<lb n="07.01°" id="NOTTEIB07.01°"/><head
id="NOTTEI">I.
<lb n="07.01b" id="NOTTEIB07.01b"/>LA NOTTE</head>
<div3 type="sequenza">
<p>
<lb n="07.01" id="NOTTEIB07.01"/>
<milestone unit="notte.1.1"/>Ricordo una vecchia
citt&agrave;, rossa di mura
<lb n="07.02" id="NOTTEIB07.02"/>e turrata, arsa su
la pianura sterminata nel-
<lb n="07.03" id="NOTTEIB07.03"/>l'Agosto torrido,
con il lontano refrigerio di
<lb n="07.04" id="NOTTEIB07.04"/>colline verdi e
molli sullo sfondo. Archi enor-
<lb n="07.05" id="NOTTEIB07.05"/>
<milestone unit="notte.1.5"/>memente vuoti di ponti
sul fiume impaludato
<lb n="07.06" id="NOTTEIB07.06"/>in magre
stagnazioni plumbee: sagome nere
<lb n="07.07" id="NOTTEIB07.07"/>di zingari mobili
e silenziose sulla riva: tra
<lb n="07.08" id="NOTTEIB07.08"/>il barbaglio
lontano di un canneto lontane
<lb n="07.09" id="NOTTEIB07.09"/>forme ignude di
adolescenti e il profilo e la
<lb n="07.10" id="NOTTEIB07.10"/>
<milestone unit="notte.1.10"/>barba giudaica di un
vecchio: e a un tratto
<lb n="07.11" id="NOTTEIB07.11"/>dal mezzo
dell'acqua morta le zingare e un
<lb n="07.12" id="NOTTEIB07.12"/>canto, da la
palude afona una nenia primor-
<lb n="07.13" id="NOTTEIB07.13"/>diale monotona e
irritante: e del tempo fu
<lb n="07.14" id="NOTTEIB07.14"/>sospeso il corso.
</p>
</div3>
<pb n="008"/>
```

Gli elementi *milestone* possono essere usati in coppie virtuali per segnalare i confini di segmenti arbitrari di testo che si sovrappongono agli elementi standard, e insieme a elementi di congiunzione o stand-off markup possono rappresentare virtualmente ogni genere di sovrapposizione, auto-sovrapposizione e segmentazione non contigua. Ad esempio la TEI consiglia questa strategia per il trattamento di segmenti varianti complesse, come le varianti del passo del Principe proposto poco sopra:

```

lasciato parte della sua gente
<anchor id=p1/>alla
<anchor id=p2/>defesa
<anchor id=p3/>della
<anchor id=p4/>obsidione
<anchor id=p5/>, con le
<-- ... -->
<-- altrove nel file -->
<app from=p1 to=p4>
<lemma wit="G L P W M U">alla difesa della</>
<rdg wit="D">alla offesa delle</>
</app>
<app from=p2 to=p5>
<lemma wit="G L P W M U">defesa della obsidione</>
<rdg wit="R.Bla">difesa di quella</>
</app>
<app from=p1 to=p3>
<lemma wit="G L P W M U">alla difesa</>
<rdg wit="B">difesa</>
</app>
<app from=p1 to=p5>
<lemma wit="G L P W M U">alla difesa della
obsidione</>
<rdg wit="A E">allobsidione</>
</app>

```

Gli elementi *milestone* offrono la massima flessibilità sintattica senza costringere a separare markup e contenuto. Per contro hanno una scarsa possibilità di essere elaborati in modo automatico mediante applicazione XML standard per finalità appena più complesse della formattazione. Infatti un *parser* XML può validare la corretta collocazione di un elemento vuoto rispetto a un modello di contenuto, o verificare che due elementi vuoti siano stati correlati mediante coppie di

attributi ID/IDREF, ma non può in alcun modo attribuire funzione strutturale alla sequenza di caratteri contenuta tra due elementi vuoti. Tale sequenza, per quanto vada considerata dal punto di vista logico un oggetto testuale non è accessibile come tale a una applicazione XML standard (e dunque anche a processori XSLT/XPATH o interpreti XQuery).

### **Oltre XML così come lo conosciamo**

Le soluzioni al trattamento delle gerarchie sovrapposte viste finora assumono come criterio centrale la compatibilità sintattica con XML sebbene, come si è rilevato, solo parzialmente questa conformità si estende anche sul piano della validazione strutturale e della possibilità di effettuare elaborazioni complesse mediante software conformi allo standard o a linguaggi basati strettamente sul modello di dati XML (come XSLT, XPATH e XQuery). In generale tutte le soluzioni proposte finora hanno grossi limiti per quel che concerne la facilità di gestione e manutenzione. Si tratta insomma di artifici sintattici che non risolvono formalmente il problema bensì, potremmo dire, lo eludono.

Per questo negli ultimi anni la ricerca teorica sullo sviluppo di sistemi di markup non gerarchici ha avuto un notevole stimolo. Allo stato non esiste nessuna proposta che si possa ragionevolmente dire risolutiva, e che rispetti tutti i criteri di adeguatezza proposti da DeRose. L'ostacolo maggiore consiste nella individuazione di un modello di dati e di un formalismo a esso associato che possa essere validato ed elaborato mediante algoritmi generali e computazionalmente trattabili come avviene per il modello ad albero di XML. Non possiamo in questa sede approfondire con il dovuto dettaglio le possibili soluzioni proposte negli ultimi anni per affrontare e risolvere in modo radicali il problema delle OH. Ci limiteremo dunque a ricordare le soluzioni che presentano maggiore interesse.

#### *XConcur*

Si tratta di un formalismo proposto da Oliver Schonefeld e Andreas Witt [2006] della Bielefeld University che eredita un costrutto già presente in SGML (CONCUR appunto) per la rappresentazione di alberi concorrenti in un medesimo documento XML. Si noti che ogni livello di markup deve essere ben-formato e di conseguenza può essere vali-

dato rispetto a uno schema o a una DTD. Ecco un esempio che riposta un frammento di codifica in sintassi XConcur:

```
<?xconcur version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE (l1)div SYSTEM "tei/dtd/teispok2.dtd">
<!DOCTYPE (l2)text SYSTEM "tei/dtd/teiana2.dtd">
<(l1)div type="dialog" org="uniform">
  <(l2)text>
    <(l1)u who="Peter">
      <(l2)s>Hey Paul!</ (l2)s>
      <(l2)s>Would you give me
    </ (l1)u>
    <(l1)u who="Paul">
      the hammer?</ (l2)s>
    </ (l1)u>
  </ (l2)text>
</ (l1)div>
```

Il limite maggiore è che non esistono allo stato sistemi per esprimere mediante questo formalismo vincoli e relazioni tra livelli gerarchici concorrenti (sebbene ci siano lavori in corso in questa direzione).

### *Just in Time Trees, JIITs*

Il sistema JIITs, proposto da Patrick Durusau e Matthew O'Donnell [2004], presenta diverse somiglianze con XConcur, poiché prevede la possibilità di esprimere alberi concorrenti in un medesimo documento. Ma introduce l'idea che un documento non debba essere ben formato fino al momento della validazione. Di conseguenza la forma serializzata di un documento può contenere markup che si sovrappone in modo arbitrario. Nel momento della validazione un filtro governato da regole seleziona l'insieme di tag che individuano uno dei possibili alberi rappresentati nel documento e solo dopo lo sottopone a controllo di validità.

Il modello di dati JIITs pertanto non si differenzia molto da quello standard di XML, sebbene un documento JIITs nella sua forma sintattica può non essere ben formato nel senso in cui lo deve essere uno XML. Proprio per questo però esso non è in grado di rappresentare complessi fenomeni di auto-sovrapposizione e non possiede alcun meccanismo per specificare relazioni e restrizioni tra i diversi livelli gerarchici.

*Layered Markup and Annotation Language (LMNL)*

LMNL (<http://www.lmnl.net>) è un modello di dati, prima ancora che un formalismo di codifica, proposto da Wendell Piez, Jeni Tennison [2002] e Paul Caton [2005]. Sono disponibili due sintassi per serializzare questo modello: una standard e una proposta recentemente da Steve DeRose e basata interamente su elementi XML vuoti opportunamente caratterizzati mediante attributi, denominata CLIX (*Canonical LMNL in XML*), che ha il vantaggio di produrre documenti XML ben formati.

In LMNL la soluzione del problema OH viene trovata uscendo definitivamente fuori dal paradigma gerarchico di XML. In questo modo è possibile rappresentare senza alcuna limitazione sovrapposizioni di qualsiasi tipo e complessità. I concetti fondamentali del *data model* di LMNL sono *range* e *layer*. I *range* sono sequenze arbitrarie di caratteri incluse tra coppie bilanciate di tag che possono sovrapporsi liberamente:

```
[1]{w}One{w} [w]is{w} [w]be-{1}
[1]fore{w} [w]two{w}.{1}
```

I *layer* sono i livelli gerarchici a cui appartengono un insieme di *range* o di *layer* subordinati. Anche essi possono sovrapporsi liberamente.

Allo stato LMNL non dispone ancora di efficaci strumenti di validazione, e tanto meno di elaborazione, e la sua radicale divergenza dal paradigma XML ha limitato molto l'interesse verso questa soluzione, sebbene i suoi ideatori siano ancora al lavoro per portarne avanti lo sviluppo.

*TexMECS e GODDAD*

TexMECS e GODDAD [Huitfeldt, Sperberg-McQueen, 2004 e 2006] rappresentano probabilmente il più avanzato e completo tentativo di trovare una soluzione generale consistente e completa al problema della rappresentazione di strutture complesse nella codifica di documenti testuali, e quello che presenta la maggiore probabilità di avere una applicazione pratica nel prossimo futuro.

Lo sviluppo di questa tecnologia è condotto nell'ambito del progetto di ricerca *Markup Languages for Complex Documents* (MLCD),

diretto da Claus Huitfeldt della University of Bergen, e con la collaborazione di Michael Sperberg-McQueen (attualmente al WorldWide-Web Consortium dopo anni di direzione della TEI), e di ricercatori della Graduate School of Library and Information Science (GSLIS) presso la University of Illinois a Urbana-Champaign.

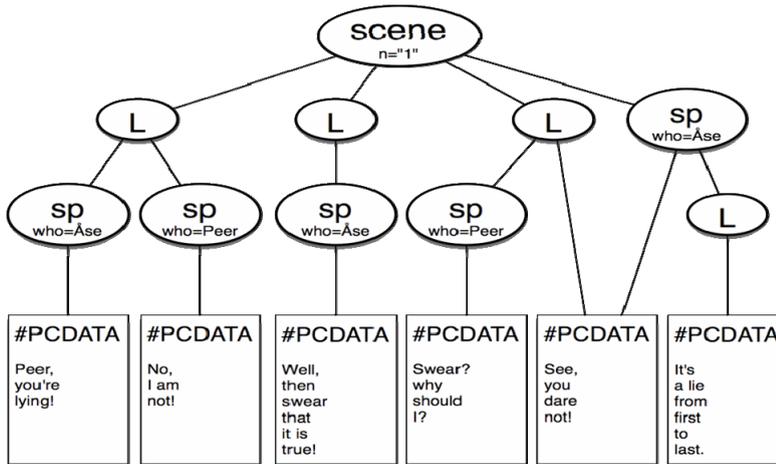
Essa rappresenta l'evoluzione delle ricerche condotte a Bergen per la creazione del *Wittgenstein Archive* (l'edizione digitale delle carte del filosofo viennese) e in particolare della formalizzazione di un linguaggio di markup abbastanza espressivo per effettuare la complessa trascrizione dei manoscritti wittgensteiniani, battezzato MECS. A differenza degli altri formalismi visti finora, MECS è stato infatti utilizzato per la creazione di un grande archivio testuale, sebbene per lungo tempo non avesse alcuna struttura dati formale sottostante e dunque non disponesse di strumenti per la validazione come avviene per XML.

Il progetto MLCD (<http://teksttek.aksis.uib.no/projects/mlcd>), ha in primo luogo potenziato la notazione originaria sviluppando TexMECS (*Trivially Extended MECS*), una notazione non molto dissimile da XML (provvede infatti anche strutture come gli attributi) la quale tuttavia permette di esprimere facilmente strutture sovrapposte, auto-sovrapposte e non contigue, come mostra il seguente esempio:

```
<act|<scene|
  <sp who="Åse" |
    <L|Peer, you're lying!|sp>
  <sp who="Peer"|<stage|without stopping.|stage>
    No, I am not!|L>|sp>
  <sp who="Åse"|<L|Well then, swear that
    it is true!|L>|sp>
  <sp who="Peer"|<L|Swear? Why should I?|sp>
  <sp who="Åse"|See, you dare not!|L>
    <L|It's a lie from first to last.|L>|sp>
|scene>|act>
```

Ma ancora più rilevante è stata la definizione di un modello di dati formale e lo studio di grammatiche formali dotate di algoritmi di *parsing* trattabili per modello di dati. Il modello di dati è stato battezzato GODDAD (*Generalized Ordered-Descendant Directed Acyclic Graph*). Si tratta di un grafo orientato acicilo dotato delle seguenti caratteristiche:

- gli archi esprimono relazione padre/figlio
- i nodi figli di ogni nodo padre sono ordinati
- i nodi padre possono condividere i medesimi nodi figli, possono cioè avere una discendenza multipla



In linea teorica questo grafo può esprimere tutte le possibili relazioni tra oggetti testuali linearizzati sottoforma di stringhe di caratteri etichettate mediante markup, inclusi i più complessi casi di auto-sovrapposizione o di frammentazione non contigua e non linearmente ordinata. Esistono tuttavia ancora dei problemi per ricavare il grafo semanticamente corretto a partire da complesse espressioni sintattiche TexMECS che ammettono diverse formalizzazioni logiche.

L'ultimo tassello, la ricerca di una grammatica equivalente ai grafi GODDAD è in corso di avanzamento, come testimoniato dalla recente presentazione di Sperberg-McQueen alla conferenza *Extreme Markup Languages* del novembre 2006. Sono stati individuati diversi candidati (*Duck-rabbit grammar*, o validazione basata su vincoli espressi in sottoinsiemi completi di logica dei predicati), il cui dettaglio tecnico tuttavia esula dagli scopi della nostra trattazione.

## Conclusioni

La creazione di risorse testuali e linguistiche su supporto informatico condotta con finalità scientifiche e culturali ha ormai una storia ultradecennale. Il grande patrimonio di risorse creato rappresenta probabilmente il successo di gran lunga più importante della “disciplina inesistente” che ci ostiniamo chiamare Informatica umanistica.

Presupposto e prodotto di questa vasta attività pratica è stata la continua riflessione circa i migliori metodi e strumenti formali per condurre il delicato compito di rappresentare quegli oggetti complessi, plurali e multiformi che sono i testi, soprattutto quelli che rientrano nella difficilmente definibile categoria dei testi letterari. Strumenti come XML e la TEI, nonostante i numerosi e acclamati limiti, hanno rappresentato e rappresentano ancora il migliore compromesso tra le complesse esigenze di adeguatezza della rappresentazione, quelle altrettanto ostiche della necessità di garantire la preservazione e la portabilità delle risorse e, ultime ma non ultime, quelle più prosaiche ma ineludibili della sostenibilità economica e organizzativa dei progetti di digitalizzazione.

Tuttavia molto rimane ancora da fare. La ricerca di soluzioni più avanzate per fare meglio ciò che è stato fatto finora è all’ordine del giorno, e come abbiamo visto le possibilità aperte sono numerose e interessanti. A questo compito è chiamata tutta la vasta e variegata comunità delle scienze del testo. Anche coloro e soprattutto coloro che finora hanno avuto una posizione critica nei confronti della svolta digitale, poiché come scrive Jerome McGann:

One of the great task lying ahead is the critical and editorial reconstitution of our inherited cultural archive in digital forms. We need to learn to do this because we don’t as yet know how. Furthermore, we scholars need to learn because it is going to be done, if not by us the by others. We are the natural heirs to this task because it is we who know most about books. [McGann, 2001:184]

## Modelli e metodi di analisi computazionale del testo

L'applicazione di metodologie informatiche nel dominio della critica letteraria presuppone la definizione di un metodo formale di analisi testuale che, attraverso livelli finiti di approssimazione, sia traducibile in uno o più procedimenti computazionali di trattamento del testo. Solo dopo aver superato questa fase si può passare alla verifica sperimentale dei procedimenti sviluppati e alla loro applicazione su un corpus di testi letterari (per ritornare eventualmente sui passaggi precedenti, allo scopo di effettuare modifiche del sistema informatico, o dello stesso quadro teorico).

Nel dominio delle discipline letterarie l'applicazione di metodi formali viene considerata con notevole sospetto. Innegabilmente la *letterarietà* o la *poeticità* di un testo dipende da fenomeni non riducibili a una descrizione formale; e tuttavia non tutto in letteratura è ascrivibile all'ineffabile. Scopo dell'analisi computazionale del testo è appunto indagare il limite dell'ineffabilità, individuare strutture e regolarità descrivibili formalmente. Come De Mauro metteva bene in evidenza alcuni anni fa, riferendosi alla ricerca in ambito linguistico:

L'informatica ci sta permettendo di sottrarre la discussione ai rischi di una contrapposizione puramente teorica di argomenti e controargomenti. A parte la ovvia portata applicativa e pratica, le tecnologie della lingua ci stanno mostrando quanta parte del comportamento linguistico umano produttivo e ricettivo sia linearizzabile, sequenziale e riconducibile ad algoritmi e quanto, invece, si affidi all'improvvisazione imprevedibile, allo scavalco di ogni regolarità. [De Mauro, 1992: 205]

L'utilizzazione delle tecnologie e delle metodologie informatiche ci impone di esplicitare il complesso di nozioni implicite nel dominio degli studi linguistici e letterari, e di sottoporle a verifica sperimentale. Essa ha dunque uno statuto teoretico diverso dal momento *critico*, che non può essere delegato alla macchina, poiché mette in gioco la soggettività e la storicità del *lettore-critico*. In ultima analisi, possiamo assumere che i limiti dell'automazione coincidano con il limite a cui si accennava sopra, il limite oltre il quale il computer (attualmente) non può andare, perché in quell'oltre si apre lo spazio dell'interpretazione.

I risultati raggiunti negli ultimi decenni dalla semiotica e dalla linguistica testuale nella comprensione dei fenomeni linguistici hanno dato un importante contributo teorico allo studio del testo letterario. Tuttavia, la prospettiva di un trattamento automatico del testo letterario che parta direttamente dal livello discorsivo, dalla catena dei significanti verbali, e possa risalire fino alla struttura semantica complessiva del testo è al di fuori della portata delle attuali tecnologie. Insomma la reale difficoltà nel trattamento automatico dei testi si presenta qualora si voglia assumere una prospettiva non quantitativa:

Sempre, quando si tenta di applicare le metodologie informatiche all'analisi del testo narrativo ci si trova di fronte a un bivio. O ci si accontenta di inseguire la parola (intesa come sequenza di caratteri alfanumerici racchiusa da due spazi bianchi) fermandosi a ciò che viene definito *pattern matching*, oppure ci si confronta con l'apparentemente irriducibile spazio che separa la materialità della pagina scritta (che è l'unica cosa che abbiamo fisicamente in mano) da tutto quello che viene compreso sotto l'etichetta di struttura narrativa. [Gigliozzi, 1991: 73]

È questo spazio tra la sequenza discorsiva dei significanti e l'universo semantico e sintattico della narrazione a essere difficilmente valicabile. Se si parte dal testo nella sua manifestazione lineare, dal livello discorsivo, il computer può fornire una grande massa di analisi quantitative e statistiche, sulla base delle sequenze di caratteri alfanumerici che rappresentano il testo nella sua memoria. Questo è, infatti, l'ambito in cui l'applicazione dell'informatica allo studio dei testi letterari ha avuto il maggiore sviluppo, con la produzione di spogli lessicali, indici di frequenze, concordanze, attività che un elaboratore elettronico svolge con una efficienza di gran lunga superiore al più volenteroso studioso, e con una estensione pressoché illimitata.

Se, invece, ci si propone di studiare i fenomeni testuali dell'intreccio, delle strutture attanziali, dei campi semantici sui quali si basa la narrazione, solo un preventivo intervento interpretativo dello studioso può fornire i dati a un sistema informatico che sia in grado di analizzare, ad esempio, la rappresentazione di un personaggio come paradigma di marche semantiche, e di individuare le relazioni possibili

tra i diversi attori della narrazione, o le sue funzioni attanziali nella struttura narrativa profonda<sup>1</sup>.

Le difficoltà nel trattamento “semantico” del testo sono legate ovviamente ai limiti degli attuali sistemi di gestione del linguaggio naturale. La comprensione sintattica e semantica di un enunciato verbale nasconde una inestricabile congerie di operazioni complesse che il cervello umano compie in quantità enorme e con una efficienza ineguagliabile. Malgrado gli sforzi volti alla formalizzazione del linguaggio naturale in atto da oltre un cinquantennio, con gli sviluppi sempre più rigorosi della linguistica generativa e della semiotica testuale<sup>2</sup>, si è ancora ben lontani dalla possibilità di eseguire/comprendere, attraverso lo strumento informatico, enunciati in lingua naturale di portata superiore alla frase. Ed è assai improbabile che si possa giungere, almeno a breve termine, allo sviluppo di sistemi sufficientemente potenti da gestire testi complessi e ambigui, sia dal punto di vista formale sia da quello tematico, come quelli letterari.

Queste considerazioni sembrerebbero respingere la praticabilità di un’analisi computazionale del testo letterario verso un orizzonte futuro. Se questo è in parte vero, si possono comunque ottenere risultati non irrilevanti attraverso lo sviluppo di procedimenti analitici computazionali che abbiano a oggetto alcuni aspetti parzialmente formalizzabili del “meccanismo” testuale, opportunamente affiancate da metodologie non computazionali. Si tratta in sostanza di rinunciare, per il momento, a un paradigma completamente computazionale e informatico, per assumere la prospettiva di un’analisi del testo (letterario) *assistita dal computer*.

In questa prospettiva, l’analisi del testo attraverso l’uso di strumenti e metodologie informatiche può essere vista come un processo che si articola in distinte fasi successive:

---

<sup>1</sup> Si vedano Gigliozzi, Giuliani, Sensini, 1987, Gigliozzi, 1991 e Gigliozzi, Giuliani, 1993a e 1993b, Giuliani, Sensini, 2002.

<sup>2</sup> Gli esempi più avanzati di queste teorie del testo sono arrivati a livelli di formalizzazione assai elevata, attraverso l’incontro della linguistica generativa con la logica modale. Si vedano gli sviluppi delle ricerche testuali di Theun van Dijk, e di János Petőfi, la cui teoria, da lui denominata TeSWeST, (da Textstruktur-Weltstruktur-Theorie), rappresenta la punta più avanzata nel tentativo di descrivere formalmente le regole di funzionamento di un testo in linguaggio naturale. Su questo si vedano Conte, 1988 e Coveri, 1984.

- 1) elaborazione di un quadro teorico di riferimento entro cui collocare i procedimenti analitici;
- 2) definizione di un modello di rappresentazione informatica o codifica del testo e delle strutture rilevanti in relazione al contesto di riferimento;
- 3) individuazione di metodi e processi di analisi testuale applicabili al modello del testo e loro definizione sottoforma di procedure formali o algoritmi;
- 4) implementazione del modello di rappresentazione e dei processi di analisi mediante adeguati linguaggi informatici;
- 5) applicazione delle procedure informatiche al testo digitalizzato;
- 6) analisi e interpretazione critica dei risultati.

Si tratta, insomma, di costruire un *modello* informatico del testo nel quadro di un determinato *contesto teorico*, per poi interrogare opportunamente tale modello e avanzare ipotesi interpretative sul testo.

L'analisi informatica del testo – pur nei suoi obblighi formali – non elimina il problema dell'oggettività del testo e delle sue interpretazioni: semplicemente si deve poter descrivere tutto quello che s'è fatto e si devono poter dare le prove ciò si è fatto. Il critico tradizionale aveva come scopo principale quello di “convincere” il lettore della bontà della propria interpretazione, magari coinvolgendolo emotivamente con la qualità della propria scrittura. L'informatica ci obbliga a basare questo coinvolgimento sull'esposizione delle prove del proprio metodo.

Il testo è uno e molti contemporaneamente e l'operazione di modellizzazione con il suo continuo fiorire di modelli possibili pare prenderne atto. Il modello che s'incarna in un determinato testo elettronico non ha alcuna pretesa di essere il modello assoluto di un altro testo ed è questo il punto in cui ci si confronta con l'esigenza di “rispondere” del proprio testo [Cazalé, Mordenti, 1997], garantendo al lettore che quel testo risponde alla volontà dell'autore e all'autore della coerenza del testo restituito con il messaggio che lui intendeva trasmettere. Questo in uno scenario in cui la molteplicità dei modelli individuabili rende problematico il ritrovamento di un'autorità garante. Il fatto è che quel testo elettronico rende concreto semplicemente uno dei modelli possibili (in un certo senso equivalenti) e che questo modello è uno strumento euristico la cui validità è assicurata semplice-

mente dalla sua capacità di descrivere una determinata rappresentazione del documento di partenza. Pongo domande al modello e ottengo risposte valide sul testo.

L'operazione di codifica resta dunque un'opera d'interpretazione (Gigliozzi, 1987) che sarà tanto più "convincente" quanto più sarà accurata la rappresentazione del testo, quanto più ricco sarà il modello del testo realizzato (a questo punto pensiamo di nuovo al livello dell'intreccio, a quello della *fabula*, agli spazi, ai tempi...).

Viene così a realizzarsi, quasi naturalmente, una sorta di rapporto ciclico tra ipotesi critica e modello che si determinano e si costruiscono a vicenda. A fianco, in conseguenza (ma ovviamente in qualche misura a prescindere dalla coppia *ipotesi critica e modello*) procede la raccolta dei materiali più vari, ma, strettamente in dipendenza dall'opzione critica che si sta attuando, dovrà essere effettuata l'identificazione e la selezione degli elementi misurabili e delle loro relazioni.

L'operazione informatica, a questo punto, si colloca nel momento preciso di un ciclo in cui misurazioni, interrogazioni e i *test* più diversi devono produrre quei materiali che porteranno al raffinamento dell'ipotesi critica e alla necessità di nuovi esperimenti. Sarà possibile, in tal modo, instaurare una sorta di rapporto dialettico tra testo, dati scaturiti dall'elaborazione informatica e ipotesi dello studioso, che realmente potrebbe aiutarci a disegnare un profilo del tutto nuovo dell'operazione di critica testuale.

Si delinea così il campo d'indagine e d'azione del critico che deve muovere dall'individuazione degli elementi misurabili nel testo (che diventeranno gli elementi del modello da realizzare) e dalla loro dichiarazione, deve passare per il riconoscimento delle relazioni che intercorrono tra queste parti (che costruiranno le relazioni tra gli elementi del modello), per arrivare, infine, alla realizzazione di un modello attendibile e utile del documento da studiare, stabilendone e dichiarandone i livelli di rappresentazione. A fianco di tutto questo sta l'identificazione del metodo e delle procedure e la loro "descrizione" e "dichiarazione".



## Tempi verbali e struttura narrativa: un modello di analisi computazionale del testo narrativo

Il contesto teorico in cui si colloca il metodo di analisi strutturale del testo che proponiamo in questo capitolo è rappresentato dalla teoria linguistica di Harald Weinrich, denominata dal suo stesso autore *C-I-T Linguistik* (ove *C* sta per *Comunicazione*, *I* per *Istruzione* e *T* per *Testo*). Fondamento di tale teoria, che il linguista tedesco ha formulato tenendo presente il concetto di *gioco linguistico* del “secondo” Wittgenstein, e la teoria degli atti linguistici di Austin e Searle, è la convinzione che la comunicazione linguistica sia da considerarsi una forma di “agire comunicativo”, e che i segni della lingua fungano da “istruzione (o ingiunzione) del parlante all’ascoltatore, affinché nella pertinente situazione questi si comporti in un determinato modo” [Weinrich, 1976 (1988: 15)].

Tali istruzioni si dividono in istruzioni semantiche, veicolate dai lessemi, e istruzioni metalinguistiche, che descrivono al destinatario di un messaggio la natura del processo comunicativo in cui si trova immerso. Weinrich [Weinrich, 1964 (1978: 37 e sgg.)], individua nella *sintassi* (nel cui ambito egli include la morfologia) la sfera della lingua che assolve alla funzione di *modellizzazione* della situazione comunicativa all’interno dell’enunciato. Egli propone di allargare la nozione di *deissi* a tutti gli elementi linguistici che aiutano il lettore a orientarsi nel processo di lettura nella sequenza di segni, poiché essi indicano al destinatario “il rapporto, e dei segni l’uno con l’altro, e dei segni con gli altri partecipanti alla comunicazione” [Weinrich, 1976 (1988: 54)]. Il testo viene così a configurarsi come una “struttura determinativa” che nel corso della lettura aiuta il destinatario a ricostruire i diversi livelli semantici, e permette il successo dell’atto comunicativo dell’emittente.

I segni *deittici* assumono nel testo il ruolo di indicatori, di istruzioni per il destinatario, e nello stesso tempo si co-determinano vicendevolmente. In questa classe rientrano tutti i morfemi ricorrenti, anaforici e temporali, quali pronomi, articoli, tempi verbali e avverbi temporali, avverbi locativi. Questi elementi morfosintattici, che hanno all’interno di un testo una ricorrenza molto elevata, sono denominati

da Weinrich “*segni ostinati*”. Essi, all’interno del processo comunicativo, rispecchiano la primitiva “preselezione del mondo” in emittente, destinatario e mondo che fonda ogni atto comunicativo:

Il processo della comunicazione, reso possibile dal codice linguistico, compare esso stesso in questo codice, precisamente con la funzione preferenziale dell’uso ostinato. [Weinrich, 1964 (1978: 15 e sgg.)]

Nell’ambito delle indicazioni teoriche di Harald Weinrich, assai vaste peraltro, lo studio sui tempi verbali rappresenta uno dei contributi fondamentali e più validi per l’analisi del testo letterario. La sua teoria dei tempi si fonda su due presupposti. Il primo è legato alla prospettiva metodologica *testuale* della linguistica di Weinrich. Vista la necessità di superare la dimensione frastica dell’analisi, ne consegue che una corretta teoria dei tempi verbali può essere enunciata solo se si considera come oggetto di studio il *testo* come entità complessa:

Nella nostra ricerca sulla problematica dei tempi verbali è indispensabile, per esempio, fissare una meta da poter descrivere all’incirca come frammento di grammatica. Ma in questo pezzo di grammatica bisogna far presente che le forme temporali [...] si incontrano soprattutto nei testi, dove essi insieme con altri segni, e quindi con altri tempi verbali, formano un complesso di determinazioni o valori testuali. Nel preparare una grammatica dei tempi questi valori testuali non bisognerà lasciarli per strada, ma piuttosto integrarli nella paradigmatica temporale, per esempio sotto forma di valori testuali indicativi. [Weinrich, 1964 (1978: 13)]

Il secondo presupposto deriva dalla critica radicale alla linguistica e alla grammatica classica, che considerano i tempi verbali come correlativi linguistici del tempo “reale”<sup>1</sup>. Per Weinrich i tempi verbali non possono essere considerati degli strumenti per esprimere il “passato”, il “presente” e il “futuro” in quanto dimensioni del tempo reale. A tale visione ontologista della lingua lo studioso tedesco oppone una teoria strutturale funzionalista dei tempi verbali, che, in quanto morfemi ostinati (presenti in gran copia, dunque, in ogni tipo di testo), fanno parte dei segni istruzionali a disposizione dell’emittente per orientare

---

<sup>1</sup> In ciò supportati dalla indistinzione nella designazione dei due oggetti nelle lingue di derivazione latina (a differenza di inglese e tedesco dove si hanno le opposizioni “tense/time” o “tempus/zeit”), la cui grammatica ha rappresentato la base degli studi linguistici per molti secoli.

la ricezione del messaggio da lui emesso. I tempi verbali sono dei “valori testuali indicativi” e appartengono alla classe dei *segni deittici*, poiché stabiliscono una relazione tra il processo di comunicazione e il testo stesso. Essi si dispongono sull’asse che lega virtualmente emittente e destinatario.

Questo processo di mediazione avviene attraverso tre funzioni fondamentali che caratterizzano il sistema dei tempi verbali. Weinrich le individua in via empirica, basandosi specialmente su spogli e analisi di testi narrativi in lingue romanze, in tedesco e in inglese:

Questa teoria consiste in effetti non nel descrivere i tempi singolarmente, separati l’uno dall’altro, ma nel raggrupparli secondo punti di vista differenti, determinando quindi la funzione di tali gruppi. Questi punti di vista li abbiamo designati come tratti distintivi del sistema temporale; essi sono (...): l’atteggiamento linguistico (commentare / narrare), la prospettiva linguistica (informazione recuperata / grado zero / informazione anticipata) e la messa in rilievo (primo piano / sfondo). [Weinrich, 1964 (1978: 220).]

La radicale sovversione delle convinzioni della grammatica classica messa in atto dal linguista tedesco (risalente nella sua prima formulazione alla metà degli anni Sessanta) ha suscitato molte critiche da parte di grammatici e linguisti, i quali non condividono la completa esclusione della referenzialità temporale e dei valori aspettuativi dalla semantica dei morfemi temporali. Malgrado queste obiezioni alla loro validità teorica generale, le tesi di Weinrich hanno avuto una notevole risonanza nell’ambito dell’analisi narratologica. A questo proposito Claude Cazalé, che ha condotto ricerche simili alla nostra sul corpus novellistico pirandelliano, rileva:

... fidandomi di precedenti esperienze confermate dalla tendenza che mi è parsa dominante nella più recente narratologia, ho scelto il sistema verbale come perno centrale nella elaborazione del racconto: ricorderò Segre che afferma: “i vari atteggiamenti dello scrittore verso la materia narrativa hanno uno dei loro tramiti fondamentali nella scelta dei tempi”. Infatti se le ipotesi e le applicazioni proposte a suo tempo da Weinrich sono state seguite e anche in parte superate, le ricerche sia della linguistica testuale tradizionale e computazionale che dei semiologi convergono per definire le conseguenze precise sui

piani sintattico, logico-semantico e tematico dell'uso dei tempi<sup>2</sup>. [Cazalé, 1989: 84]

Lo stesso Weinrich, d'altra parte, ha affermato di avere sviluppato la sua teoria “osservando i problemi linguistici con l'occhio del letterato e i fenomeni letterari con l'interesse mosso dagli interrogativi della linguistica”<sup>3</sup>. Occorre notare, peraltro, che la maggior parte degli esempi utilizzati dallo stesso Weinrich per verificare ed esporre le sue tesi sui tempi verbali sono tratti da testi letterari e narrativi, e comunque nella totalità da testi scritti. Sebbene il linguista tedesco non intenda limitare le sue tesi sui tempi verbali ai testi scritti, e specificamente ai testi letterari, dal nostro punto di vista metodologico è proprio verso questo tipo di testi che le sue indicazioni risultano particolarmente produttive.

A differenza dei processi comunicativi linguistici orali, il testo narrativo, in quanto testo *scritto* ha uno statuto comunicativo peculiare. Ciò che rende la comunicazione letteraria scritta manifestamente differente dall'interazione verbale orale è il fatto che emittente (autore) e destinatario (lettore) non sono in rapporto di contiguità spaziale e temporale. La distanza spazio-temporale tra gli attori del processo comunicativo determina una radicale trasformazione della *sfera pragmatica*, ovvero della complessa rete di rapporti che si instaurano tra emittente e destinatario, tra questi e il messaggio, tra messaggio e contesto sociale e circostanziale (ad esempio è impossibile far ricorso agli usuali stratagemmi metacomunicativi dell'interazione linguistica dialogica, come il *feed-back*). In particolare gli aspetti relativi al rapporto emittente-destinatario subiscono un processo di semiotizzazione, che li trasforma in strutture formali del testo, come suggerisce Segre:

... nel caso dell'opera letteraria la linea emittente-ricevente si spezza in due parti: emittente-messaggio e, con soluzione di continuità, messaggio-emittente, ecco allora che l'emergere di atteggiamenti e interventi personali dell'autore si trasforma in dato formale, proprio del messaggio; non è

---

<sup>2</sup> La citazione di Segre è tratta da *Semiotica e filologia*, Torino, Einaudi, 1979, p. 39.

<sup>3</sup> Questa affermazione di Weinrich è contenuta nel discorso tenuto dal linguista in occasione della sua elezione a membro dell'Accademia delle Scienze della Renania-Westfalia, ed è riportata da Segre nella “Prefazione” alla traduzione italiana di Weinrich, 1976.

l'emittente che si rivolge a noi, è il messaggio che contiene, come elemento del suo artificio, le apostrofi o le meditazioni attribuite all'emittente, alla stessa stregua in cui contiene i discorsi e i pensieri dei personaggi. Ed è per il tramite di questa comunicazione a due fasi che i problemi dei tempi modi aspetti e così via si rivelano di natura squisitamente formale. [Segre, 1974: 29]

Questa osservazione ci permette di richiamare la definizione di Weinrich dei segni sintattici, in quanto segni che all'interno del testo hanno la funzione di istituire e orientare il processo comunicativo, come istruzioni che il lettore deve seguire per recepire la catena sintagmatica correttamente. La distribuzione e la sequenza dei tempi verbali nel testo, e le loro reciproche transizioni, costituiscono dunque una parte della manifestazione sul livello discorsivo del testo, come strutture formali, del rapporto comunicativo autore-lettore. Riportando le indicazioni di Weinrich nell'ambito dell'analisi del testo letterario, Cesare Segre ne individua nel campo della coerenza testuale e nello studio del punto di vista i punti di applicazione più interessanti:

Tra i molti impieghi critici di queste osservazioni sull'opposizione tra i due tipi di verbi [commentativi vs. narrativi] ne segnalo qui due: 1) la distribuzione di questi verbi segue norme abbastanza costanti, e perciò collabora sia alla coerenza del testo, sia alla segnalazione del suo inizio e della sua fine; 2) l'alternanza dei tipi di verbi contribuisce all'istituzione dei piani narrativi (per esempio il *primo piano* e lo *sfondo*), e perciò è un elemento costitutivo della prospettiva del racconto: i vari atteggiamenti dello scrittore verso la materia narrata hanno uno dei loro trami fondamentali nella scelta dei tempi. [Segre, 1985: par. 1.6.3]

Si può ipotizzare, dunque, che i tempi verbali, attraverso la loro distribuzione paradigmatica e sintagmatica nel testo narrativo, contribuiscano a manifestare sul livello discorsivo del testo (e dunque sul piano strettamente linguistico) i seguenti aspetti strutturali della narrazione:

- 1) rapporto tra autore/narratore e materia della narrazione (eventi e stati narrati);
- 2) realizzazione sul livello discorsivo dei rapporti cronologici tra intreccio e fabula;
- 3) successione sintagmatica sul livello discorsivo delle sequenze narrative ed espansioni descrittive.

Per la precisione al punto (1) si collega l'opposizione tra tempi commentativi e tempi narrativi; al punto (2) la divisione dei tempi tra

funzione retrospettiva e funzione anticipativa; al punto (3) la divisione dei tempi narrativi tra tempi del primo piano e tempi dello sfondo.

Naturalmente la fenomenologia di questi elementi della semiotica narrativa non viene esaurita dalla distribuzione di tempi verbali, e a essa contribuiscono in misura notevole gli aspetti semantici della lingua. Si può dire che il sistema temporale dei verbi costituisce un quadro strutturale di fondo che consente al lettore di orientarsi nella ricezione del testo, uno schema rispetto al quale ogni autore costruisce le sue deviazioni idiolettali. Solo un ampio spoglio, allargato a testi narrativi e non, ed eventualmente a testimonianze della lingua orale di una certa epoca, potrebbe fornire risultati più precisi su queste funzioni della lingua<sup>4</sup>.

### **L'atteggiamento linguistico e il punto di vista**

La prima funzione testuale attribuita da Weinrich al sistema dei tempi verbali nel testo è l'espressione dell'*atteggiamento linguistico* dell'emittente nei confronti del materiale linguistico. In base a tale funzione i tempi verbali si dividono tra *tempi narrativi* e *tempi commentativi*:

... i valori indicativi del commentare e del narrare, inerenti in qualità di tratti distintivi ai morfemi temporali con ricorrenza ostinata offrono a chi parla la possibilità di influire in una maniera determinata sull'ascoltatore guidandolo nell'atto ricettivo di un testo. Il parlante infatti, usando i tempi commentativi dà a capire che per lui è opportuno che l'ascoltatore nel recepire quel tal testo assuma un atteggiamento di *tensione*, mentre coi tempi narrativi dà ad intendere, per opposizione, che il testo può essere recepito in stato di *distensione*. [Weinrich, 1964 (1978: 44)]

Possiamo individuare in questa opposizione uno dei tramite che manifesta sul livello discorsivo del testo la struttura dei punti di vista narrativi, ovvero i fenomeni legati al modo della narrazione.

---

<sup>4</sup> Una interessante applicazione su un corpus abbastanza rilevante dello schema di Weinrich alla lingua italiana è lo studio di Cristina Lavinio relativo a testi fiabeschi sia scritti sia orali [Lavinio, 1984]. Altri cenni alle teorie di Weinrich nel campo della semiotica dei testi letterari, sono contenute nei saggi di Roncallo, Testa e Ravazzoli in Coveri, 1984.

La categoria di “modo della narrazione” è stata definita teoricamente da Genette, che la distingue dal fenomeno, attiguo ma differente, della “voce” [Genette, 1972]<sup>5</sup>. Mentre la voce specifica il tipo di presenza dell’istanza narrativa nella narrazione, che può essere interna o esterna rispetto alla narrazione, e delle istanze enunciative dei personaggi, il modo, come scrive Segre, è “legato al rapporto scrittore-materia narrata” [Segre, 1981: 100].

In particolare, all’interno del livello diegetico la successione di tempi del “mondo narrato” e di tempi del “mondo commentato” consente di distinguere le parti diegetiche in senso stretto, il racconto di eventi e stati, dalle parti in cui l’istanza narrativa si intromette direttamente nel testo per manifestare il suo pensiero e le sue opinioni da una posizione esterna alla narrazione, con una funzione che potremmo definire *metanarrativa*. Questa alternanza è particolarmente evidente nelle narrazioni con narratore eterodiegetico, ovvero esterno alla storia, ove l’alternanza tra tempi narrativi (quasi sempre alla terza persona singolare, la persona verbale più frequente nella narrazione) e tempi commentativi segnala gli interventi metanarrativi del *narratore onnisciente*.

Nella narrazione mimetica, invece, ove sono riportati i discorsi e i dialoghi dei personaggi, i tempi verbali segnalano i punti di vista dei personaggi. Infatti questi ultimi sono i responsabili di un atto enunciativo fittizio di secondo grado, in relazione al quale sono organizzati i tempi, solo in seconda istanza mediati dall’istanza narrativa autoriale<sup>6</sup>. Le parti dialogate, dunque, possono a loro volta *raccontare* o *commentare* gli eventi del livello narrativo primario, ma dipendono da una enunciazione secondaria, che a differenza dell’istanza narrativa primaria è una enunciazione intradiegetica. I tempi si dispongono nelle parti mimetiche in relazione a questa situazione comunicativa.

---

<sup>5</sup>. Lo studioso francese con la sua sistemazione ha posto ordine a una superfetazione concettuale e terminologica che una tradizione quasi secolare di ricerche aveva portato nello studio di questi fenomeni. Si veda per una rassegna su questo argomento Segre, 1981, e anche Segre, 1984: 15-28 e 124-131.

<sup>6</sup> Genette ha fatto una distinzione simile, limitando il suo discorso, però, solo al fenomeno della *narrazione nella narrazione*: “Definiremo la differenza di livello dicendo che *ogni avvenimento raccontato da un racconto si trova ad un livello diegetico immediatamente superiore a quello dove si situa l’atto narrativo produttore di tale racconto*”. [Genette, 1972 (1976: 275)].

Questa distinzione permette di rendere più complessa la tipologia dell'atteggiamento linguistico dominante in un testo, tenendo presente la rilevanza delle parti mimetiche e la distribuzione dei tempi al loro interno, ovvero lo statuto comunicativo del testo che si muove tra i due opposti poli della *mimesi* nel dramma e nel romanzo epistolare puro, e della *diegesi* nel romanzo privo di discorso diretto. Peraltro si può presumere che ogni genere narrativo, e all'interno di ciascuno di esso, ogni poetica narrativa, sia caratterizzata da una certa distribuzione standard dei tempi verbali, che dipende in parte dall'evoluzione del sistema-lingua nel suo complesso, in parte da regole specifiche di sintassi narrativa. Questa indagine si può spingere fino alla ricerca di uniformità e variazioni idiolettali ascrivibili a un certo autore, in modo da costruire una sorta di diagramma evolutivo di schemi nell'uso di tempi verbali.

La distribuzione dei tempi verbali tra tempi commentativi e tempi narrativi dipende dal sistema verbale di ciascuna lingua. Weinrich prende in esame nel suo libro il tedesco, l'inglese, il francese, lo spagnolo e l'italiano. La struttura delle lingue neolatine è molto simile, e dunque i tempi per l'italiano sono divisi in nei due gruppi come segue:

<b>Tempi commentativi</b>	<b>Tempi narrativi</b>
Presente	Imperfetto
Futuro	Perfetto
Futuro anteriore	Piucchepperfetto I
Passato prossimo	Piucchepperfetto II
	Condizionale presente
	Condizionale passato

Si deve rilevare che questa ripartizione dei tempi verbali dell'italiano moderno non va presa come assoluta. Il passato prossimo, infatti, nella comunicazione orale quotidiana sostituisce molto spesso il perfetto, che tende a scomparire dall'uso linguistico dei parlanti. Uno studio effettuato da Cristina Lavinio sull'uso dei tempi verbali nelle fiabe orali in confronto con quelle scritte, ha peraltro rilevato come la stessa narrazione orale faccia prevalere in generale i tempi commentativi piuttosto che i tempi narrativi [Lavinio, 1984]. Viceversa lo schema resta valido per le novelle scritte, specialmente quelle rielaborate o riportate da letterati o scrittori veri e propri. In effetti si

può affermare che solo nei testi scritti il perfetto mantiene la sua funzione di tempo principale della narrazione.

### La prospettiva linguistica e gli universi temporali nella narrazione

La seconda funzione testuale attribuita da Weinrich ai tempi verbali è la prospettiva linguistica. Essa è determinata dallo svolgimento necessariamente lineare di un enunciato linguistico, che rende inevitabile lo sviluppo temporale della emissione/ricezione della sequenza di segni:

... l'insieme dei segni di un testo [...] ha nella catena di segni (*chaîne parlée*) una disposizione lineare. Questo è il flusso dell'informazione [...]. Ogni segno linguistico ha pertanto nel testo un prima e un dopo testuale, e sia la preinformazione che la postinformazione concorrono a determinarlo. In generale possiamo quindi parlare di tempo testuale. [Weinrich, 1964 (1978: 77)]

Il concetto di tempo testuale è strettamente connesso con la fenomenologia della lettura<sup>7</sup>. Weinrich infatti rileva come “sarebbe [...] errato presumere che alla linearità della catena di segni corrisponda anche un progredire lineare da segno a segno dell'operazione decodificatrice” [Weinrich, 1964 (1978: 77)]. Ancora una volta una osservazione di Cesare Segre ci sembra assai pertinente per determinare gli aspetti fondamentali di tale fenomenologia:

... il lettore di un libro legge, di volta in volta in volta una sola frase; tutte le precedenti vengono a costituire una sintesi memoriale (di contenuti, di elementi stilistici, di suggestioni), mentre quelle ancora da leggere formano un'area di possibilità sia linguistiche, sia narrative. [Segre, 1974: 16]

La lettura dunque procede in una duplice dimensione, una legata allo svolgimento lineare del testo, l'altra caratterizzata da continui slittamenti tra la *riattualizzazione* di ciò che è stato recepito precedente-

---

<sup>7</sup> In realtà si possono estendere queste osservazioni anche alla ricezione di enunciati orali, tenendo presente però quanto si diceva apertura di capitolo sulla diversità del contesto pragmatico che caratterizza l'interazione dialogica orale dalla comunicazione scritta. Un esempio di ricezione di un enunciato orale che si avvicina al meccanismo della lettura è la comunicazione teatrale, fatte salve le innovazioni di certi autori d'avanguardia che hanno portato gli spettatori *realmente* dentro il dramma.

mente, e l'attesa previsionale verso ciò che il lettore presume debba seguire. Il linguista tedesco parla a questo proposito di *preinformazione* e *postinformazione* rilevando che la ricezione è basata su tale movimento complesso a partire dalla decodifica del livello linguistico del testo. Questo processo è regolato e orientato da alcuni segnali testuali che hanno una funzione di "deissi intratestuale", o deissi interna: in senso anaforico, determinando la riattualizzazione di informazioni già date; o in senso cataforico, spingendo l'attenzione e l'attesa del destinatario verso i nuovi contenuti che saranno comunicati nel seguito del testo<sup>8</sup>. Il sistema dei tempi verbali contribuisce a costruire questo sistema di indici intratestuali:

Il tempo testuale insieme con le sue direzioni fondamentali relative alla preinformazione e alla postinformazione lo incontriamo di nuovo nel sistema dei tempi verbali. Oltre alla possibilità di regolare l'atteggiamento linguistico secondo i gruppi di tempi verbali del mondo commentato e del mondo narrato, il sistema temporale offre delle differenziazioni, che consentono di orientarsi in relazione al tempo testuale e in particolare danno modo di disporre liberamente entro certi limiti. Vale a dire che in questo modo un'informazione può essere recuperata o anticipata. [Weinrich, 1964 (1978: 78)]

La prospettiva linguistica serve a esprimere la relazione tra il *tempo testuale* e il "tempo reale". Con questa locuzione si intende "il momento o il decorso del contenuto della comunicazione" [Weinrich, 1964 (1978: 78)]. Se il rapporto tra il tempo testuale e il tempo reale non ha sfasamenti la prospettiva linguistica viene definita da Weinrich di *grado zero*; se viceversa tale rapporto comporta un recupero o una anticipazione di informazioni rispetto al decorso delle ricezione testuale si è in presenza rispettivamente di una *retrospezione* e di una *previsione*. Entrambi i gruppi temporali principali, tempi del "mondo commentato" e tempi del "mondo narrato", si dividono al loro interno relativamente alla funzione di manifestare la prospettiva linguistica secondo il seguente schema:

---

<sup>8</sup> Oltre ai tempi verbali hanno una funzione determinante nella struttura anaforica e cataforica del testo gli articoli, appunto divisi in determinativi e indeterminativi, e gli aggettivi/pronomi determinativi, che svolgono tale funzione attraverso la loro divisione tra definiti e indefiniti. Si veda sull'articolo francese in particolare Weinrich, 1974.

	<b>Commento</b>	<b>Racconto</b>
<b>Grado zero</b>	Presente	Perfetto Imperfetto
<b>Retrospezione</b>	Passato prossimo	Piuccheperfetti
<b>Previsione</b>	Futuri	Condizionali

Attraverso la prospettiva linguistica, il sistema dei tempi verbali nell'ambito di una narrazione contribuisce a realizzare linguisticamente, sul livello discorsivo (cui è legato l'andamento temporale della lettura), la struttura temporale dell'intreccio<sup>9</sup>, e dunque la sequenza degli universi temporali della narrazione. In generale si può dire che ai tempi della retrospezione sul piano narrativo corrispondono le analessi narrative, ovvero i *flash-back*, mentre ai tempi della previsione corrispondono le prolessi o anticipazioni.

Questo schema generale necessita di alcuni approfondimenti. In primo luogo, affinché vi sia una vera e propria analessi narrativa, occorre che sia presente nel testo una sequenza cospicua di tempi verbali retrospettivi, tali da costituire delle sezioni sufficientemente vaste della narrazione. In secondo luogo una analessi o una prolessi può essere retta da una istanza enunciativa di secondo grado, ad esempio nelle sezioni mimetiche del racconto. In questo caso tale analessi viene espressa o dai tempi retrospettivi del mondo commentato, o dai tempi narrativi *tout-court*, a seconda dal punto di vista del personaggio.

Infine si deve osservare che le distinzioni di Weinrich sulla funzione dei tempi verbali, come per la divisione tra tempi commentativi e tempi narrativi, necessitano di una revisione che va misurata sulle particolarità linguistiche e letterarie di una data epoca come di un determinato autore. A questo proposito Cazalé, nel suo studio sulle novelle di Pirandello, rileva come "il sistema tradizionale dei tempi si complica" [Cazalé, 1989: 85]. Infatti nella narrativa pirandelliana il condizionale e l'imperfetto, insieme a inserzioni di forme verbali al presen-

---

<sup>9</sup> L'intreccio è una parafrasi riassuntiva del testo in cui gli eventi sono disposti nella stessa sequenza in cui si trovano nel livello discorsivo. Si veda anche Segre, 1974 e 1985.

te, assumono spesso la funzione di segnalare il monologo interiore o il soliloquio dei personaggi, e dunque non sono propriamente indirizzati verso il futuro (anche quello fittizio del mondo narrato), quanto piuttosto verso un universo temporale del “non mai realizzato”. Questa osservazione esemplifica la variabilità diacronica e idioletale del sistema funzionale dei tempi verbali nella struttura narrativa, variabilità della quale l’analisi strutturale deve tenere sempre conto.

### **Il rilievo narrativo nel testo**

L’ultima opposizione funzionale che caratterizza il sistema dei tempi verbali è ricondotta da Weinrich al *rilievo narrativo*. Questa funzione riguarda la distinzione tra i due tempi narrativi con prospettiva di grado zero, in quelle lingue (quali tutte le lingue neolatine) in cui tale distinzione viene marcata linguisticamente. Il linguista tedesco rileva come l’alternanza tra questi due tempi verbali (passato remoto e imperfetto nell’italiano) dà “*rilievo* a una narrazione articolandola in senso ricorrente in primo piano e sfondo” [Weinrich, 1964 (1978: 128)]. Attraverso il concetto di rilievo narrativo Weinrich espunge dalla semantica dei tempi verbali le nozioni di *aspetto* e *azione*, in base alle quali la grammatica rende conto dell’opposizione tra imperfetto e passato remoto (denominato appunto perfetto in relazione alla sua natura aspettuale perfetta). A questa idea del linguista tedesco si sono opposti molti linguisti, che ritengono errato abbandonare le categorie di aspetto e azione<sup>10</sup>. Senza approfondire queste tesi in maniera specifica, occorre rilevare che lo stesso Weinrich, pur essendo critico sulla statuto teorico dei concetti di “aspetto” e “Aktionsart” verbali, afferma:

...osserviamo innanzitutto, per motivi di metodo, che nel corso di queste considerazioni non ci occuperemo più dell’“aspetto” o della “natura dell’azione” o di cose del genere. Questi concetti, checché essi possano voler significare secondo i singoli studiosi, si riferiscono alla frase. [Weinrich 1964 (1978: 128)]

---

<sup>10</sup> Si veda ad esempio Bertinetto, 1986.

In questo modo il linguista tedesco non esclude del tutto l'aspetto dalla descrizione dei tempi verbali, ma lo relega alla grammatica della frase, rivendicando una specificità per la grammatica dei testi. La nozione di piano narrativo, come lo stesso Weinrich rileva, non è però facilmente definibile, a meno di una argomentazione circolare per la quale "è sfondo tutto ciò che sta all'imparfait [imperfetto in italiano] e primo piano tutto ciò che sta al *passé simple* [passato remoto]" [Weinrich, 1964 (1978: 128)]. Weinrich ricorre pertanto a una definizione alquanto impressionistica e contenutistica, definendo primo piano "ciò per cui una storia si racconta", il "*fatto inaudito*".

Nell'ambito della narratologia e della semiotica letteraria con la metafora di "piano", ripresa dal mondo cinematografico, si suole indicare una serie di fenomeni non sempre omogenei<sup>11</sup>. La nozione di rilievo narrativo di Weinrich si colloca sul livello del discorso, ovvero sulle modalità linguistiche attraverso le quali l'intreccio si realizza come sequenza di enunciati linguistici. Si può dire che la funzione del rilievo narrativo dei tempi verbali manifesta al livello linguistico la successione dei motivi narrativi principali con quelli secondari. La successione dei tempi verbali narrativi determina la segmentazione del livello discorsivo in sequenze e costituisce un elemento del ritmo narrativo, che l'autore può dosare diversamente. Le sequenze al passato remoto, tempo del primo piano, indicano sul livello discorsivo del testo gli snodi dell'intreccio, gli eventi o azioni, mentre le sequenze discorsive all'imperfetto sono tipiche delle espansioni descrittive.

Anche qui occorre rilevare che i concetti di "azione" e "descrizione" vanno sempre intesi in rapporto all'universo semantico del testo, alla poetica narrativa dell'autore, e in ultima istanza agli orizzonti di attesa di una data cultura. A questo proposito è condivisibile la topologia delle descrizioni proposta di Agostino Roncallo nell'ambito di un vasto spoglio della tradizione novellistica italiana:

... alla [descrizione] possono essere assegnate due funzioni: una "dilatatoria", per cui nel testo si viene a creare un "ritardo", una attesa che arresta la progressione della storia, e una "demarcativa-organizzatrice" che tende, invece, a

---

<sup>11</sup> Una definizione del concetto di piano narrativo è data da Lotman, 1970 (1990: 306). Ci sembra, però, che questa definizione, sebbene sicuramente appropriata, faccia rientrare il "piano narrativo", nell'ambito dei fenomeni narrativi che regolano il rapporto tra il narratore e lo spazio-tempo fisionale del racconto.

sottolineare le articolazioni della narrazione e ad inquadrare l'enunciato propriamente narrativo, assicurando così la concatenazione logica e, spesso, sillogistica, dello stesso enunciato (una specie di *post hoc ergo propter hoc*). [Roncallo, 1984: 155.]

Weinrich stesso, assumendo le vesti del critico letterario piuttosto che del linguista, evidenzia come la scelta dei tempi nell'*incipit* e nell'*explicit* della narrazione, e la dominanza complessiva dei tempi di un tipo piuttosto che di un altro sono elementi dello stile narrativo, e dunque vanno riportati alla poetica dominante in una certa epoca, e a quella specifica di un autore. Interessanti a questo scopo le sue osservazioni sulla narrativa naturalista, nel cui ambito la prevalenza dei tempi dello sfondo rispetto ai tempi del primo piano va ricollegata all'importanza che la poetica naturalista attribuisce alla definizione dell'ambiente sociologico e storico in cui i protagonisti vivono le loro vicende.

### **Le transizioni temporali e la coerenza testuale**

Chiudiamo questo inquadramento teorico-metodologico dell'analisi dei tempi verbali in ambito narratologico con alcuni cenni relativi alle *transizioni temporali*. Il termine di transizione, che Weinrich mutua dalla *teoria dell'informazione*, denota "il passaggio da un segno all'altro nel corso dello svolgimento lineare del testo" [Weinrich, 1964 (1978: 217)]. Considerando il testo verbale come una sequenza ordinata di segni (lessemi e morfemi), all'inizio del processo comunicativo il destinatario del messaggio si trova in uno stato di informazione zero (massima entropia), poiché tutti i segni del codice possono apparire con il medesimo indice di probabilità (la fonte è in uno stato di "equiprobabilità", nei termini della teoria dell'informazione). Il processo della ricezione si può rappresentare come una progressiva "riduzione delle probabilità" da parte del ricevente, dunque un aumento della quantità di informazione in suo possesso.

Naturalmente l'applicazione del modello matematico informazionale al testo linguistico come struttura significativa è improponibile. Per rendere produttive queste osservazioni in un ambito linguistico (e tanto più di semiotica del testo letterario) occorre rinunciare a una loro applicazione in termini puramente matematici, e limitare il campo di osservazione a un genere particolare e limitato di segni, con una se-

mantica non ambigua, come appunto il sistema dei tempi verbali nella teoria di Weinrich. In questo modo il concetto di transizione assume una valenza anche semiotica. Se nella sequenza del testo all'inizio tutte le probabilità sono aperte, con il procedere della ricezione alcuni stati divengono più probabili di altri, e il destinatario del messaggio può prevedere con una certa sicurezza il segno che seguirà.

Nel campo dei morfemi temporali del testo, questo significa che il lettore, con il procedere della lettura si costruisce un sistema di attese a livello grammaticale, che vengono di volta in volta confermate o non confermate. In generale, data la comparsa di un certo morfema, è più probabile che il morfema seguente sia funzionalmente omogeneo, e una contravvenzione a questa regola probabilistica costituisce un segnale notevole all'interno del testo. Sulla scorta di questa osservazione Weinrich rileva:

Qualunque sia la categoria sintattica presa in esame, nei testi le transizioni più frequenti sono di regola quelle omogenee [...]. Nel quadro di una teoria linguistica testuale questo fenomeno trova senz'altro una spiegazione. Le transizioni omogenee garantiscono la consistenza di un testo, la sua *testualità* [...]. Ciò significa che in un testo a un massimo di transizioni omogenee corrisponde un massimo di testualità. [Weinrich, 1964: (1978: 225)]

In questo modo il sistema dei tempi verbali contribuisce alla costituzione della *coerenza del testo*, come rileva Segre. Infatti nella ricezione del testo il lettore si aspetta che, una volta apparso un certo morfema temporale, il successivo appartenga alla medesima categoria funzionale, ovvero che la transizione sia omogenea.

Il sistema dei tempi verbali, in altri termini, costituisce una *isotopia grammaticale*<sup>12</sup>, che si sviluppa lungo l'asse sintagmatico del testo connettendo in *paradigma* elementi testuali distanti. Come il testo è

---

<sup>12</sup> Il concetto di *isotopia* come noto è stato introdotto da Greimas, e indica "l'iteratività lungo una catena sintagmatica di classemi che assicurano al discorso-enunciato la sua omogeneità" [Greimas, Courtés, 1979: 197-199]. Il concetto, che in origine si riferiva solo al livello semantico di un testo è stato poi esteso anche al livello dell'espressione, poiché le ricorrenze di alcuni elementi morfosintattici e fonetici costituiscono dei livelli di senso autonomi da quello veicolato dal livello semantico primario del testo verbale (su questo illuminante Lotman, 1970). Una interessante rielaborazione del concetto di isotopia nel contesto di una teoria semiotica dell'atto di lettura viene avanzata in Eco, 1979: 93-101.

pluri-isotopo al livello semantico, poiché ogni lessema ha marche semantiche che lo legano a diversi percorsi di senso, così anche i tempi verbali si dispongono nel testo su diverse isotopie.

Ogni morfema temporale costituisce isotopie grammaticali con tutti i tempi del testo che svolgono la stessa funzione, entrando in relazione con tre diverse catene isotopiche (corrispondenti alle tre marche distintive del sistema dei tempi verbali). Nel testo si vengono in tal modo a creare diverse isotopie, che il lettore durante il processo di lettura scopre progressivamente. La costruzione di isotopie temporanee costituisce un elemento fondamentale del processo di lettura e di comprensione del testo. Se la transizione da un morfema temporale al successivo è omogenea (ovvero il tempo verbale seguente è dello stesso tipo del precedente) l'isotopia costruita dal lettore viene confermata. Viceversa il lettore deve modificare il suo sistema di attese.

Nel testo considerato lungo il suo sviluppo sintagmatico si possono avere diversi tipi di transizioni temporali eterogenee, in relazione alla complessità del sistema verbale di una data lingua. Weinrich distingue tra transizioni di primo grado, che vedono il passaggio transizionale di una sola dimensione funzionale, da quelle di secondo e terzo grado, che egli denomina *metafore temporali*. Mentre le transizioni di primo grado costituiscono con il loro succedersi i segni del passaggio da una isotopia temporale a un'altra, le metafore temporali assumono un valore "controdeterminante" o attenuativo rispetto al contesto in cui sono collocate, e dunque non hanno una funzione fissa.

La distribuzione e la quantità delle transizioni temporali, dunque, fornisce uno strumento ulteriore per l'analisi testuale attraverso lo studio dei tempi verbali. Il testo letterario si colloca, anche dal punto di vista dei morfemi temporali, lungo una scala che va dalla massima prevedibilità, alla massima apertura. Naturalmente, alla costituzione delle isotopie, degli orizzonti di attesa, contribuiscono le regole che pertengono sia al genere letterario sia ai canoni retorici di una certa fase diacronica della storia letteraria. In base a questo sostrato codificato, ciascun autore, o ciascun testo, se si preferisce, si qualifica per innovazione, o per conferma della tradizione.

## Conclusioni

Il quadro teorico che abbiamo esposto nelle pagine precedenti, a partire dai suggerimenti di Weinrich, dimostra come il sistema dei

tempi verbali rappresenti il tramite per introdursi nella struttura di un testo narrativo. E poiché i morfemi temporali sono elementi linguistici che si collocano immediatamente sul livello discorsivo del testo, essi sono accessibili a un modello di analisi computazionale che non richieda l'elaborazione automatica di strutture semantiche "profonde". In questo contesto la costruzione del modello informatico del testo cui si accennava in apertura consiste nella *rappresentazione*, mediante un opportuno linguaggio di codifica, della sua struttura componenziale (suddivisione in unità strutturali e separazione tra segmenti diegetici e segmenti mimetici) e dei morfemi temporali.

Sugli aspetti teorici e tecnici della rappresentazione digitale del testo esiste ormai una ingente mole di riflessioni. Basti ricordare come, allo stato attuale, le soluzioni più avanzate in questo settore siano rappresentate dagli schemi di codifica basati su linguaggi di *mark-up* SGML/XML, in particolare quello sviluppato dalla *Text Encoding Initiative* (TEI). Questi schemi di codifica consentono di rappresentare esplicitamente le caratteristiche di un testo mediante marcatori la cui funzione e applicazione è regolata da una grammatica formale.

Un testo codificato con questo genere di linguaggi di *mark-up* può essere interrogato mediante uno dei numerosi strumenti di *information retrieval* XML/SGML attualmente disponibili<sup>13</sup> al fine di ottenere dati dettagliati circa la distribuzione sintagmatica e paradigmatica dei morfemi temporali. Tali dati quantitativi debbono infine essere esaminati per analizzare sia la struttura paradigmatica dei fenomeni legati al sistema dei tempi verbali, sia la disposizione sintagmatica delle medesime categorie verbali nel testo (cioè, lungo lo sviluppo del *livello discorsivo*), al fine di studiarne la struttura narrativa.

---

<sup>13</sup> In questa sede era nostro intento spiegare in modo dettagliato i procedimenti tecnici e gli strumenti informatici che possono essere adottati per effettuare queste analisi, quanto piuttosto indicare i problemi teorici dell'analisi computazionale del testo e soffermarci sul metodo di analisi narratologica da noi proposto. Per coloro che sono interessati ai dettagli tecnici rimandiamo ai numerosi siti Web dedicati a XML e in particolare al sito *XML Cover Pages*, <http://www.oasis-open.org/cover>.



## ***Il turno di Pirandello: un esperimento di analisi assistita dal computer***

Questo capitolo descrive i risultati di un esperimento di analisi del *Turno* di Luigi Pirandello ‘assistita dal computer’<sup>1</sup>.

Nonostante gli importanti sviluppi delle tecnologie informatiche, l’applicazione di metodi computazionali nell’analisi dei testi letterari pone ancora numerose difficoltà. Allo stato attuale, infatti, non disponiamo di sistemi affidabili che consentano di estrarre automaticamente informazione semantica da un testo in lingua naturale, meno che mai da un testo letterario. Di fatto gli strumenti di analisi testuale si basano ancora oggi su processi di analisi quantitativa e di elaborazione statistica<sup>2</sup>. Tuttavia l’adozione di opportune strategie di codifica testuale consente l’applicazione di questi processi su unità testuali più complesse delle semplici occorrenze delle forme verbali presenti nel testo. Il linguaggio di codifica con cui si costruisce la rappresentazione digitale del testo, se sufficientemente potente, diventa un linguaggio teorico mediante il quale si possono esplicitare in modo formale caratteri-

---

<sup>1</sup> Due motivi ci hanno indotto a scegliere nel *corpus* pirandelliano questo romanzo. Il primo è di tipo strettamente strumentale. L’estensione (tra romanzo e racconto lungo) di quest’opera si presta alla sperimentazione delle tecniche di elaborazione informatica “pionieristiche” adottate nella ricerca. Il secondo, più propriamente critico-letterario, è la collocazione liminare del *Turno* nell’ambito del processo di disgregazione del romanzo verista e naturalista attuato da Luigi Pirandello. Questo consente di analizzare le modalità linguistiche e strutturali di tale decostruzione e ristrutturazione interna della struttura narrativa naturalista in una delle sue prime testimonianze. Nel corso dell’esposizione i brani del romanzo vengono citati dall’edizione critica inclusa in *Tutti i romanzi (Tutti i romanzi)*, a cura di Giovanni Macchia con la collaborazione di Mario Costanzo, coll. “I Meridiani”, Mondadori, Milano 1990), alla quale si rinvia con “*Il turno*” e il numero di pagina. Per i saggi e gli scritti teorici, abbiamo utilizzato la raccolta *Saggi, poesie, scritti varii* (a cura di Manlio Lo Vecchio Musti, Mondadori, Milano 1960), alla quale si rinvia con “*Saggi*” e il numero di pagina in cui si trova l’articolo citato.

<sup>2</sup> Su questo si veda Gigliozzi, 2003 e Hockey, 2000.

stiche e strutture testuali astratte, che divengono in questo modo disponibili all'analisi.

Il compito che si pone allo studioso, pertanto, è quello di elaborare un quadro teorico entro cui iscrivere l'analisi, dal quale deriva un modello di rappresentazione del testo che deve essere formalizzabile mediante il linguaggio di codifica. È questo modello che va sottoposto a indagine automatica, mediante opportune procedure informatiche di interrogazione; i risultati dell'indagine a loro volta, debbono diventare oggetto del lavoro interpretativo. In questo senso abbiamo adottato la prudente formula di analisi 'assistita dal computer': la macchina informatica, infatti, diventa uno strumento analitico che si integra nel metodo critico, ne determina il carattere formale, ma non lo esaurisce.

Muovendo da queste considerazioni metodologiche, il nostro lavoro ha richiesto in prima istanza l'individuazione del quadro teorico di riferimento. Da questo punto di vista ci è sembrato che lo studio del sistema dei tempi verbali secondo le indicazioni a suo tempo fornite da Harald Weinrich [Weinrich, 1978] potesse fornire importanti indicazioni per l'analisi narratologica, come numerosi studiosi hanno rilevato<sup>3</sup>. Abbiamo già in altra sede esposto i dettagli di questo contesto teorico [Ciotti, 2002]. Qui ci limiteremo pertanto a ricordare che i tempi verbali, attraverso la loro distribuzione paradigmatica e sintagmatica nel testo narrativo, contribuiscono a manifestare sul livello discorsivo del testo alcuni importanti aspetti strutturali del testo narrativo:

- 1) rapporto tra autore/narratore e materia della narrazione (eventi e stati narrati);
- 2) realizzazione discorsiva dei rapporti tra intreccio e fabula;
- 3) articolazione sintagmatica delle sequenze narrative e descrittive.

Per la precisione al punto (1) si collega l'opposizione suggerita da Weinrich tra *tempi commentativi* e *tempi narrativi*; al punto (2) la divisione dei tempi tra *funzione retrospettiva* e *funzione anticipativa*; al

---

<sup>3</sup> Oltre al già citato Weinrich, 1978, si vedano Segre, 1974: 29 e segg., Segre, 1985: par. 1.6.3 e Cazalé, 1989.

punto (3) la divisione dei tempi narrativi tra tempi del *primo piano* e tempi dello *sfondo*.

Sulla base di questo quadro teorico si è proceduto alla codifica informatica del testo pirandelliano. A questo fine abbiamo adottato il linguaggio di mark-up XML sviluppato dalla *Text Encoding Initiative*, che ci ha consentito di rappresentare in modo formale le caratteristiche strutturali e linguistiche del testo rilevanti ai fini dell'analisi<sup>4</sup>. Oltre alla struttura editoriale (suddivisione in capitoli e capoversi), sono stati codificati esplicitamente i morfemi verbali del testo, identificandone i tempi, e le sequenze in discorso diretto. L'interrogazione del testo è avvenuta mediante le funzionalità di *information retrieval* del software *Dynatext*, una potente applicazione in grado di gestire documenti XML. Questo strumento, a differenza dei comuni sistemi di ricerca *full text*, permette di effettuare ricerche contestuali (ovvero di selezionare le sezioni del testo entro cui svolgere la ricerca) sia per stringhe di testo sia per occorrenze di determinati elementi XML<sup>5</sup>.

Tale modalità di ricerca ha consentito di estrarre i dati relativi alla distribuzione e alla frequenza dei morfemi temporali nelle varie sezioni del *Turno*, distinguendo opportunamente tra livello diegetico e livello dialogico.

### **Tempi verbali e configurazione enunciativa del romanzo**

L'indagine, secondo lo schema di analisi semiotica del testo poetico tracciato da Lotman [1990], si è divisa in due momenti:

- 1) analisi della distribuzione paradigmatica dei tempi in tutto il testo, nel tentativo di individuare alcune modalità idiolettali dell'autore nell'uso del sistema temporale;
- 2) analisi della disposizione sintagmatica delle medesime categorie verbali (cioè, lungo lo sviluppo del livello discorsivo del testo), al fine di ricostruire la struttura compositiva del testo.

---

<sup>4</sup> Su XML e *Text Encoding Initiative* rimandiamo a Ciotti (a c. di), 2003 e Sperberg-McQueen, Burnard (a c. di), 2001.

<sup>5</sup> In questa sede non era nostro intento spiegare in modo dettagliato i procedimenti tecnici e gli strumenti informatici adottati nella nostra ricerca.

*Il turno* per estensione può essere considerato un racconto lungo piuttosto che come un vero e proprio romanzo. Il testo si divide in trenta brevi capitoletti, alcuni non più lunghi di una pagina, i quali spezzano, quasi fino alla frammentazione, la sequenza della narrazione. Questa frammentazione del discorso narrativo, che peraltro si ritrova anche in altre prove narrative dello scrittore siciliano, permette di allungare il “tempo della storia” in confronto al “tempo del discorso”, pur senza appesantire la narrazione. I capitoletti in cui è suddiviso il racconto non sono titolati, e hanno una numerazione ordinale progressiva che, rileva Guglielminetti, non supera “il livello di un segno grafico di sospensione indispensabile nel corso della lettura” e risponde a un criterio di partizione della narrazione “ancora dentro le formule compositive del romanzo zoliano” [Guglielminetti, 1986: 75-76].

La tipologia dell'enunciazione narrativa resta legata ai canoni della *impersonalità*. Il narratore, rigidamente extradiegetico, conduce infatti tutta la narrazione in terza persona. Questa struttura del testo veicola un contenuto narrativo che sembra collocare il romanzo nella sfera di influenza del regionalismo verista siciliano. La rassegna dei motivi narrativi presenta gli elementi tipici di una rappresentazione quasi sociologica delle tradizioni culturali isolane, più volte rappresentate da Pirandello in novelle e romanzi della sua prima fase narrativa: l'istituzione matrimoniale, con le sue formalità, la gelosia maschile, accompagnata spesso da un inveterato gallismo, la posizione subordinata della donna, come figlia e come moglie, costituiscono tutti elementi tematici tipici della narrativa veristica di ambiente siciliano.

L'analisi dei dati sulla distribuzione dei morfemi temporali, come vedremo, se per molti aspetti ha confermato la collocazione del romanzo in questo contesto letterario, ha permesso di evidenziare numerosi indizi di un processo di ‘decostruzione dall'interno’ della struttura narrativa naturalista.

I morfemi temporali complessivamente registrati nel testo sono 3139, che, su un totale di 26731 parole rappresentano lo 11,2%<sup>6</sup>. I dati

---

<sup>6</sup> Il rapporto percentuale tra tempi verbali e totale delle parole in un testo non rappresenta una semplice curiosità statistica, poiché i tempi verbali rappresentano in un certo modo lo scheletro dell'impianto propriamente narrativo di un racconto, i segni verbali che ne veicolano l'essenza di esposizione di eventi e azioni. Weinrich per ricavare l'indice di “ostinazione” dei tempi verbali si riferisce al rapporto tra quantità di tempi e numero di righe a stampa. Nel nostro caso avremmo, malgrado un leggero

relativi alle frequenze temporali nell'intero testo indicano che il 57% dei tempi è composto da tempi narrativi, mentre il 43% da tempi commentativi. La predominanza dei tempi narrativi sui commentativi si presenta dunque molto limitata, rispetto a quanto non ci si dovrebbe aspettare in un racconto con un forte impianto verista e una configurazione enunciativa totalmente extradiegetica<sup>7</sup>. La peculiarità di questa distribuzione complessiva dei tempi verbali viene confermata dal fatto che il tempo verbale con frequenza maggiore è il presente, il principale del livello commentativo, che ha una frequenza relativa del 34%, mentre i due tempi narrativi fondamentali, perfetto e imperfetto, incidono entrambi per il 24% sul totale.

La distribuzione della massa linguistica del testo tra parti diegetiche (narrazione pura) e parti mimetiche (discorso diretto) fornisce dati rilevanti ai fini dell'analisi testuale e dell'interpretazione di questi dati. Infatti, la caratteristica più rilevante nella struttura del testo è costituita dalla rilevanza delle parti mimetiche: come si è giustamente scritto, "nel romanzo domina il dialogo, un dialogo potenzialmente teatrale, che investe il lettore e che va pronunciato" [Angelini, 1976: 368]. Questa rilevanza del livello enunciativo intradiegetico viene attestata assai bene da una ricognizione statistica sulla distribuzione delle parole nel testo. Il testo consta infatti nella sua totalità di 26731 forme verbali, delle quali 10790 sono in sezioni mimetiche e 15941 in sezioni diegetiche. L'incidenza percentuale del discorso diretto è dunque del 40%.

Sulla scorta di queste osservazioni si può dunque ipotizzare che la struttura fortemente drammatizzata della narrazione determini un pro-

---

scarto, una conferma delle sue osservazioni che fissano la distribuzione media dei tempi verbali pari a uno per linea di testo a stampa. Si tratta di un indicatore privo di fondamenti linguistici e statistici, ma se ne può tenere conto in prima istanza. Solo una quantità di rilevamenti molto vasta potrebbe far scaturire risultati meno impressionistici da questi dati.

<sup>7</sup> Si confrontino questi dati percentuali con le rilevazioni condotte dallo stesso Weinrich, che rileva nei testi narrativi una frequenza dei tempi narrativi intorno all'80%. Anche i dati di Cristina Lavinio per le fiabe di Capuana indicano una percentuale dei tempi narrativi che oscilla dall'86% al 97%. Purtroppo si tratta di dati non omogenei, e per di più condotti su testi brevi, che dunque hanno una struttura diversa dal punto di vista della distribuzione delle sezioni mimetiche, per cui il confronto risulta solo indicativo, ma comunque servono a evidenziare la distorsione nella distribuzione delle frequenze temporali che presenta *Il turno*.

cesso di distorsione nella distribuzione relativa delle frequenze temporali tra le varie classi funzionali. Una conferma all'ipotesi avanzata sopra viene dall'analisi separata dei dati di frequenza temporale tra mimesi e diegesi. I tempi verbali infatti si dispongono in maniera fortemente orientata tra i due livelli enunciativi. Nelle parti diegetiche vi è una assoluta predominanza dei tempi narrativi (99,6% del totale), mentre i tempi commentativi sono del tutto assenti, con sole sette occorrenze del presente; nelle parti mimetiche si concentrano al contrario la maggior parte dei tempi commentativi.

Si evidenziano perciò due "modalità enunciative" predominanti: una prima, in cui il narratore extradiegetico esercita il controllo sull'enunciazione, caratterizzata dai tempi narrativi; una seconda, in cui l'autore lascia la responsabilità dell'enunciazione ai personaggi stessi, caratterizzata dai tempi commentativi. L'analisi delle transizioni temporali conferma che il romanzo si dispone su due livelli di enunciazione ben distinti, che costituiscono due isotopie temporali. Si rilevano quattro picchi di frequenza: i due massimi si hanno in corrispondenza delle transizioni omogenee dell'atteggiamento linguistico (commento/commento e racconto/racconto); a questi corrispondono simmetricamente le alte frequenze delle transizioni eterogenee tra tempi narrativi e tempi commentativi. L'esigenza di alternare questi due livelli determina la forte incidenza di transizioni eterogenee tra tempi del commento e tempi del racconto, presenti in misura assai maggiore di quanto non ci si aspetterebbe in un testo narrativo.

Se passiamo a esaminare i tempi nelle parti mimetiche, esse risultano, come ci si aspetta, connotate dalla predominanza dei tempi commentativi (88% del totale), tra cui il presente da solo costituisce il 70%, mentre i tempi narrativi sono poco rilevanti, malgrado non del tutto assenti.

Dunque il dialogo dei personaggi si riferisce prevalentemente in senso commentativo ai fatti dell'intreccio. Le sequenze diegetiche di secondo grado, come attesta la scarsa presenza di tempi narrativi nel livello mimetico, sono molto limitate: piuttosto che narrare infatti, i personaggi del romanzo attualizzano i fatti e gli eventi del passato, utilizzando il tempo retrospettivo del commento, il passato prossimo. Tuttavia il sistema temporale del dialogo si presenta complessivamente più variato rispetto a quello del livello diegetico, presentando con percentuali rilevanti tutti i tempi verbali (con l'eccezione dei piuccheperfetti).

Le frequenze delle transizioni dei tempi commentativi anticipativi e retrospettivi con il presente sono più alte di quelle omogenee, a ulteriore conferma di un tessuto verbale meno coerente e compatto del livello diegetico puro. Altrettanto rilevante nelle parti mimetiche la frequenza del condizionale I, che dal punto di vista transizionale determina l'occorrenza di molte transizioni di secondo grado, con valore di metafora temporale attenuativa [Weinrich, 1978: 255 e seg.].

La distribuzione temporale relativa al livello enunciativo diegetico, con la quasi totale assenza di tempi commentativi (le poche forme verbali al presente nel livello diegetico risultano tutte forme di presente descrittivo) ribadisce la configurazione extradiegetica del romanzo. La voce narrante non interviene mai direttamente nel corso della narrazione, come istanza soggettiva, autonoma dai personaggi, per commentare i fatti dell'intreccio e le motivazioni dei vari personaggi, o per partecipare delle loro vicende.

Il discorso diretto invece presenta due tipologie enunciative. Accanto a sequenze di scene dialogate introdotte dai *verba dicendi*, segnalate dalle transizioni tra tempi narrativi e tempi commentativi, si collocano intere sequenze di lunghezza variabile al "discorso diretto libero". Si veda il seguente esempio:

- Non capisco, - disse Ciro, infoscandosi. - Per prepotenza?
- Per prepotenza... eh già! Capisci? sono in cinque...
- E tu, babbeo? Hai detto che non saresti andato?
- Nient'affatto!
- Ma intanto sei qua... E hai paura! Te lo leggo negli occhi: hai paura! Ah, ma tu ci andrai, stasera stessa, or ora... Prepotenze, neanche Dio! Vieni con me.
- Dove?
- In casa Alcozèr!
- Ora?
- Ora stesso. Il tempo di vestirmi. A che ora suoli andarci tu?
- Alle sei e mezzo. [*Il turno*: 269-270]

In queste sequenze, in cui Pirandello dimostra peraltro una eccezionale maestria nel comporre un vero e proprio dialogo comico, arriva a compimento il processo di dissoluzione dell'impianto narrativo: la scomparsa dell'istanza diegetica, che lascia il posto a una configurazione totalmente drammatica, corrispondente a quella della scrittura teatrale, determina la dissoluzione della realtà oggettiva, che viene ormai rifratta solo attraverso la prospettiva dei personaggi, padroni assoluti della scena.

### **La prospettiva linguistica: temporalità progressiva dell'intreccio e irrazionalità degli eventi.**

La scelta dei tempi costituisce un mezzo a disposizione dell'autore per segnalare sul livello discorsivo del testo le *acronie* che intervengono tra gli eventi attualmente narrati e la loro sequenza cronologica effettiva, che il lettore ricostruisce con il procedere dell'atto di lettura. L'analisi dei tempi verbali consente, dunque, di esaminare i rapporti oppositivi tra l'intreccio e la *fabula*.

La distribuzione temporale nel *Turno* indica una sostanziale linearità cronologica della narrazione. Nelle sezioni diegetiche si ha una predominanza pressoché completa dei due tempi con prospettiva di grado zero, che costituiscono il 90% del totale. I tempi retrospettivi sono il 7%, mentre i tempi anticipativi sono appena il 3% (vedremo che il condizionale, tempo dell'anticipazione ha in Pirandello una funzione particolare). La linearità compositiva del romanzo trova una conferma nella quasi assoluta supremazia delle transizioni omogenee tra i tempi del primo piano.

Le sequenze diegetiche retrospettive con una dimensione non irrilevante sono solo due: una di carattere eterodiegetico, riporta uno squarcio del passato di donna Bettina, la madre di Pepè [*Il turno*: 238-239]<sup>8</sup>. La seconda invece è di carattere omodiegetico, e connette i capitoli VIII e IX, con una breve narrazione relativa all'esito del duello tra Pepè e il Borrani. Una particolarità di questa analessi è la progressiva modificazione della prospettiva narrativa che dal punto di vista del narratore extradiegetico passa a quello dei due padrini, trasformandosi in un discorso indiretto libero:

Ma il Tucciarello e il D'Ambrosio non **erano** imbronciti per l'esito doloroso del duello; bensì per il contegno del loro primo di fronte all'avversario. Non che Pepè avesse fatto propriamente una cattiva figura; ma, appena impugnata la sciabola, Cristo santo! – **pensava** il Tucciarello, morsicchiandosi con le labbra la punta della barba, – appena impugnata la sciabola, **era diventato** più pallido di una carogna; per poco le braccia non gli **eran cascate** su la persona, come se la sciabola fosse stata di bronzo massiccio. Parare? sfalsare? Niente! Lì come un pupazzo da teatrino... E allora, si sa, zic-zac, al primo scontro,

---

<sup>8</sup> Per la distinzione tra analessi omodiegetiche, ovvero interne alla storia principale del racconto, e analessi eterodiegetiche, esterne rispetto alla storia principale del racconto, si veda Genette, 1976: 96 e segg.

pàffete! Meno male, che non se l'era presa in testa. Il Borrani lo avrebbe spaccato in due, come un mellone. [*Il turno*: 244-245]

Nemmeno il procedimento dell'analessi mimetica, il "racconto del personaggio", è presente in misura consistente, come si era già osservato analizzando la configurazione temporale della mimesi: l'unico caso proprio di racconto di secondo grado è la storia dei fantasmi narrata da don Diego a Pepè [*Il turno*: 302-303].

Nel *Turno*, dunque, la narrazione si articola secondo una temporalità progressiva, in cui gli eventi seguono la "freccia del tempo". L'intreccio non si oppone alla *fabula*, non presenta quella multidimensionalità temporale che caratterizzerà la narrativa pirandelliana posteriore, dove il succedersi di analessi e prolessi, insieme all'introduzione vistosa del monologo interiore, sarà funzionale alla scomposizione della realtà, realizzata tramite una "perdita di valore del tempo stesso come dimensione di ordine nello svolgersi dei fatti" [Guglielminetti, 1986: 71].

Da questo punto di vista il romanzo sembra risentire in misura maggiore dell'influenza dei canoni naturalistici. Tuttavia, anche in questo caso l'autore utilizza un procedimento narrativo di derivazione naturalistica per disarticolarlo dall'interno: la struttura cronologica sequenziale del racconto viene di fatto contraddetta e forzata dal livello tematico dell'intreccio.

Basta guardare non tanto agli eventi in sé (già si è detto che una rassegna dei motivi fornirebbe la ricetta per un perfetto raccontino regionalista), quanto ai rapporti tra essi nello svolgimento della narrazione: dove la *necessità* controlla i destini dei personaggi che popolano la narrativa naturalista e verista, già in questa precoce prova narrativa del giovane Pirandello emerge in misura prepotente l'azione del *caso*. La catena evenemenziale dell'intreccio infatti si sviluppa per una serie di casualità che frustrano ogni volta i progetti dei personaggi, e gli orizzonti d'attesa del lettore (si pensi poi al lettore di romanzi di fine Ottocento).

Paradigmatiche a questo proposito sono le vicissitudini dei due protagonisti, Marcantonio Ravi e Pepè Alletto. Il romanzo si apre proprio con la voce, al discorso diretto, del Ravi, che "va in giro per la città, fermando parenti e conoscenti" per spiegare l'impeccabile ragionamento alla base del suo progetto di maritare la figlia con l'ultrasessantenne don Diego. Ma tutti gli sforzi che egli compie per

realizzarlo producono, con variabile intervallo di tempo, e attraverso catene di eventi imprevedibili, l'effetto contrario.

Simmetrica la vicenda di Pepè Alletto: casualmente si ritrova designato come promesso secondo sposo di Stellina, e altrettanto casualmente si impegna in un duello; quando poi comincia a maturare realmente il progetto di sposarla dopo la morte del vecchio, per un ulteriore intervento del caso, fa incontrare il cognato con Stellina, e la perde. Infine, solo una suprema svolta del fato, permette a entrambi di raggiungere l'oramai insperato obiettivo. Lo scioglimento del racconto, con la morte di Ciro che lascia Stellina vedova, e sistemata, costituisce una vera e propria celebrazione dell'irrazionalità paradossale della vita<sup>9</sup>.

La disposizione cronologica progressiva degli eventi, dunque, entra in rapporto oppositivo con il livello tematico dell'intreccio producendo, anche su questo piano, una contraddizione strutturale. Laddove la sequenzialità dell'intreccio, nella narrativa naturalista, era funzionale alla narrazione di una sequenza di eventi logicamente legati da rapporti di causa-effetto – quasi, date le ipotesi iniziali, la catena degli eventi procedesse, attraverso deduzioni, verso lo scioglimento finale – nel *Turno* cronologia e logica evenemenziale si oppongono, determinando una tensione irrisolta, che porta il romanzo in un territorio, se non proprio esterno alla tradizione naturalista, almeno proprio ai limiti di frontiera.

### **Rilievo narrativo e tipologia dei procedimenti narrativi**

La ripartizione dei tempi narrativi tra primo piano e sfondo nelle parti diegetiche, presenta una situazione di sostanziale equilibrio: il passato remoto costituisce il 45% del totale e l'imperfetto il 43%. Questa equidistribuzione dei tempi tra i due piani del rilievo narrativo,

---

<sup>9</sup> Lo scioglimento del romanzo, con la morte di Ciro Coppa, pure se risponde a questa "logica del caso" messa in atto dall'Autore, sembra infatti un vero e proprio *coup de theatre*, e sorge il dubbio con Tedesco, se "si sostituisca qui all'*ethos* della commedia il *pathos* della tragedia, sia pure familiare. È il culto siciliano della famiglia e della morte, l'ebbrezza, il farnetico dell'abito nero, che conservano il loro valore fra i tanti negati valori, oppure è la scena patetica il definitivo intervento surrettiziamente dissacratorio della mente ironicamente critica?" [Tedesco, 1976: 47-48].

segnala la rilevante presenza, nelle sezioni diegetiche del romanzo, di segmenti genericamente “descrittivi”.

Le transizioni temporali nell’ambito del rilievo narrativo mostrano come i valori delle frequenze delle transizioni temporali omogenee (in totale 937) sono nettamente superiori a quelli delle transizioni eterogenee (271). All’interno del gruppo delle transizioni omogenee, vi è una ulteriore polarizzazione: 595 transizioni sfondo-sfondo a fronte di 342 primo piano-primo piano. Da questo prospetto si può inferire che entrambi i tempi narrativi si dispongono preferenzialmente in sequenze omogenee lungo lo sviluppo testuale, e che le sequenze di tempi dello sfondo, in questo ambito, sono più compatte delle sequenze al primo piano.

Un elemento interessante che emerge dall’analisi delle transizioni temporali è il forte numero di transizioni del tipo perfetto-presente (207) e viceversa (219). A questo dato di frequenza si aggiungono i risultati di un sommario spoglio lessicale delle forme verbali che mostra, nel perfetto (tempo del primo piano) la netta prevalenza dei *verba dicendi*. Su un totale di 734 occorrenze, oltre trecento sono forme di *verba dicendi*, le quali costituiscono, del resto, le forme verbali con maggiore frequenza: 51 occorrenze della forma “disse”, 28 di “gridò” e “rispose”, 21 di “esclamò”. Allargando lo spoglio a tutti i verbi al perfetto, si può rilevare come la metà dei tempi del primo piano narrativo regge, introduce o accompagna il discorso diretto legato.

Il racconto diegetico di primo piano viene dunque sostituito molto spesso da sezioni di monologo o di dialogo tra i personaggi, producendo una intensa drammatizzazione degli snodi narrativi dell’intreccio, delle azioni, che si concentrano intorno a poche sequenze sceniche.

Questa distribuzione lessicale dei verbi al perfetto aumenta il peso relativo del tempo imperfetto nel sistema temporale complessivo della diegesi. Secondo Weinrich la tendenza ad aumentare la consistenza della narrazione allo sfondo che si verifica nella narrativa europea del secolo scorso, è legata alle scelte di poetica dei narratori naturalisti di dare maggiore importanza alla descrizione sociologica, al contesto

spazio-temporale in cui si inserisce il fatto e i personaggi oggetto della narrazione<sup>10</sup>.

Tuttavia, a un esame più approfondito, *Il turno* rivela un deciso processo di rifunzionalizzazione nell'uso dell'imperfetto, che mostra *in actu* una vera e propria decostruzione dall'interno della sintassi narrativa naturalista.

Il tempo verbale dello sfondo, infatti, non sempre corrisponde all'inserzione di espansioni descrittive dilatorie, tese a rallentare la narrazione con descrizioni ambientali o, ritratti fisici dei caratteri. Le espansioni descrittive vere e proprie sono anzi molto limitate e di breve estensione<sup>11</sup>. Piuttosto prevale un tipo di espansione che descrive “non tanto le circostanze ambientali e paesaggistiche, quanto, piuttosto, comportamenti, azioni, stati d'animo” [Roncallo, 1984: 156], la cui funzione è quella di legare e giustificare gli eventi dell'intreccio:

se [...] con la descrizione “dilatoria”, ci troviamo di fronte ad una interruzione, linearmente molto estesa, della sintagmatica del racconto che impone alla lettura un momento di pausa, la descrizione di tipo demarcativo si propone all'opposto, come esponente di un “continuum” narrativo non mediato o interrotto da procedimenti intesi a rallentare il procedere della storia. [Roncallo, 1984: 155-156]

Questo tipo di espansioni descrittive sono caratterizzanti dal punto di vista del genere letterario, poiché si tratta di un procedimento narra-

---

<sup>10</sup> Si veda quanto afferma Weinrich, analizzando l'uso della messa in rilievo nel romanzo: “A partire dal XIX, da Balzac in poi, il romanzo si fa realistico, e ciò vuol dire sociologico. Gli autori non si contentano ormai di raccontare una storia più o meno bella ed emozionante, ma aspirano in misura crescente a dare informazioni attendibili – e per i naturalisti persino dal punto di vista scientifico – sulle condizioni sociali del tempo. A questo fine l'azione principale, che di solito comporta, come in precedenza, solo poche persone, viene inserita in una rappresentazione sempre più vasta dello sfondo sociologico... La relazione tra primo piano e sfondo nella narrazione viene a invertirsi nel senso che adesso lo sfondo diventa il più importante dei due e il primo piano il meno importante” [1978: 135-136].

<sup>11</sup> Pirandello infatti tratteggia luoghi e personaggi con pochi tratti, molto caustici in genere, tesi a produrre una vera e propria deformazione espressionistica. Questa caratteristica del narratore siciliano è stata individuata per la prima volta da Giacomo Debenedetti in relazione all'ultima raccolta pirandelliana di novelle *Una giornata* [Debenedetti, 1945], ma recentemente i lavori di Graziella Corsinovi hanno individuato i sintomi di una anticipazione espressionistica anche nella narrativa giovanile di Pirandello.

tivo rintracciabile prevalentemente nella novellistica piuttosto che nel romanzo e conferiscono al *Turno* un ritmo narrativo molto veloce, che funge da contrappunto ai frequenti rallentamenti dovuti alle sequenze di scene dialogate.

In questo ambito l'uso del tempo imperfetto, come tempo dello sfondo, si diversifica, determinando una ulteriore diffrazione tipologica dei procedimenti narrativi. Nella distribuzione co-testuale l'imperfetto svolge differenti funzioni testuali e narrative, a seconda del contesto morfologico e semantico in cui viene collocato, e dell'appartenenza a sequenze lineari omogenee o a sequenze lineari eterogenee con altri tempi. Nella nostra analisi abbiamo rintracciato alcune invarianti funzionali, che consentono di abbozzare un primo tentativo di tipologia dei procedimenti narrativi a esso legati:

1) Sequenze omogenee con funzione descrittiva dilatoria.

Sono le sequenze omogenee all'imperfetto che costituiscono le espansioni descrittive dello spazio narrativo (ambiente e paesaggi), o dei personaggi. Nel *Turno* vi sono, come detto, poche occorrenze di questo tipo, e hanno tutte una estensione lineare molto limitata (si veda ad esempio *Il turno*: 256: "Il cielo... qualche gazza").

2) Sequenze omogenee con funzione demarcativa-esplicativa.

Queste espansioni narrano in genere eventi ripetitivi, ma non necessariamente la natura aspettuale del verbo è legata a quella della narrazione. La loro presenza serve invece a introdurre una fase dell'intreccio al primo piano, a spiegare l'insieme di motivazioni che conduce agli eventi del momento trasformativo vero e proprio, spesso indugiando in tratteggiamenti psicologici degli stati d'animo dei personaggi, condotti tuttavia dal punto di vista esterno della voce narrante (si veda ad esempio *Il turno*: 285-286: "Stellina lo comprendeva... il suo rancore").

3) Sequenze eterogenee con i tempi commentativi con *funzione demarcativa*.

Questo particolare uso dell'imperfetto si trova nel contesto di scene di dialogo tra i personaggi collocate allo sfondo narrativo invece che al primo piano, con una tecnica compositiva tipica nella narrativa pirandelliana. Ciò spiega il numero rilevante delle transizioni di secondo grado tra imperfetto e tempi commentativi (in entrambi i casi intorno alle 120 occorrenze). Come le sequenze demarcative omogenee, han-

no la funzione di narrare eventi e azioni che introducono i momenti trasformativi dell'intreccio al primo piano. Guglielminetti ha evidenziato come l'inserzione di verbi di azione all'imperfetto tra segmenti di discorso diretto serva anche a descrivere gli atteggiamenti del personaggio che parla. Si tratta in sostanza di un uso scenico del tempo narrativo dello sfondo, che serve a tratteggiare le pose del personaggio durante l'atto di enunciazione e svolge la funzione di una didascalia nella scrittura drammaturgica<sup>12</sup>.

- 4) Sequenze omogenee con funzione *espositiva interna* del punto di vista dei personaggi sugli eventi dell'intreccio.

Si tratta delle sequenze di imperfetti che veicolano le varie tipologie di discorso indiretto libero. Vista l'importanza di queste sequenze nel sistema narrativo pirandelliano è opportuno dedicare loro una trattazione specifica e particolareggiata.

### **Discorso indiretto libero e soliloquio dei personaggi**

Nel *Turno* vi sono diverse sequenze di discorso indiretto libero marcate del punto di vista linguistico e morfologico. Risulta molto utile, a questo riguardo, una osservazione dallo stesso Weinrich.

Nella narrativa del novecento l'inserzione, nel corpo della narrazione, del punto di vista dei personaggi mediante il discorso indiretto libero deve fare a meno dei tradizionali segnali che denotano il cambiamento del soggetto dell'enunciazione e della prospettiva narrativa. In sostituzione vengono utilizzati una serie di segni sintattici e semantici: dal punto di vista morfologico, il discorso indiretto libero viene segnalato "negativamente", attraverso l'assenza dalle sequenze narrative del perfetto, mentre semanticamente si caratterizza per una accentuazione della presenza di verbi modali. Un ulteriore procedimento si pone al livello propriamente stilistico-formale, ovvero la preferenza per lo "stile orale", caratterizzata da una costruzione disarticolata del

---

<sup>12</sup> A questo proposito, una interessante analisi contrastiva delle strutture testuali di *L'altro figlio* novella con la omonima trascrizione drammaturgica, condotta da Flavia Ravazzoli [1984: 336-337] ha evidenziato come nella maggior parte dei casi questo genere di verbi imperfetti venga trasformato nel presente delle didascalie d'autore nel copione drammaturgica.

periodo, da frequenti ellissi, da interrogazioni, da frasi troncate, con forte presenza di punteggiatura [Weinrich, 1978: 230 e segg.].

Questo insieme di caratteristiche linguistiche e stilistiche si ritrova in moltissime sequenze del *Turno*, che dal punto di vista dei tempi risultano dominate dall'imperfetto, con innesti di condizionale. Il condizionale come tempo dell'anticipazione, e dunque dell'irreale, si presta infatti a essere utilizzato in contesti in cui i personaggi si lasciano prendere dalla fantasia, e danno forma linguistica ai propri desideri o rimpianti, immaginando una serie di eventi controfattuali. Si veda ad esempio questo passo del primo capitolo:

Se egli fosse stato ricco, se avesse potuto far da sé la felicità della figliuola – bella forza! si sa, non l'**avrebbe data** in moglie a quel vecchiccio. Stellina certo, per il momento, non **poteva** apprezzare la fortuna che egli le **procacciava**: questo **era** naturale e in certo qual modo scusabile! Di lì a pochi anni però – ne **era** sicuro – ella lo **avrebbe lodato, ringraziato e benedetto**. Non **sperava**, né **desiderava** nulla per sé, da quel matrimonio; lo **voleva** unicamente per lei, e **stimava** dover suo di padre, dover suo di vecchio provato e sperimentato nel mondo, tener duro e costringere la figliuola inesperta a ubbidire. Lo **amareggiava** invece profondamente la disapprovazione di uomini d'esperienza come lui. [*Il turno*: 215]

Questo passo testimonia la distanza del procedimento narrativo di Pirandello dallo *stream of consciousness* (ancora maggiore di quella che separa lo stile narrativo di Svevo da quello di Joyce, come ha rilevato Debenedetti). Infatti mai nella narrativa di Pirandello prende il sopravvento quel *discorso dell'altro* (dell'inconscio), quei contenuti mentali prerazionali che caratterizzano il monologo interiore vero e proprio [Debenedetti, 1971: 594-616]. Il monologo interiore dei personaggi pirandelliani è sempre pensiero verbalizzato, anche quando si tratta di raziocini paranoici.

Per questo ci sembra che la definizione di *soliloquio*, data da Guglielminetti sia la migliore per descrivere il carattere enunciativo dei monologhi pirandelliani nel *Turno*: si tratta di un parlare a voce alta per “sostenere in pubblico argomenti palesemente assurdi” [Guglielminetti, 1986: 63]. La parola infatti è sempre parola pronunciata ad alta voce anche nei casi, in cui sembra che la narrazione si addentri nella interiorità dei personaggi, e ciò è testimoniato dalla indifferenza nell'uso del discorso indiretto libero e del discorso diretto che si alternano vicendevolmente nel corso della narrazione.

Torniamo dunque a ritrovare come caratteristica precipua del romanzo la parola dei personaggi. Si tratta come visto di una parola tal-

volta immediata, che prende il sopravvento sulla narrazione, e si presenta direttamente al lettore in tutte le sue articolazioni; talaltra di una parola mediata dalla voce narrante, che si pone in una prospettiva interna ai personaggi, e ne articola i pensieri in un soliloquio farneticante. *Il turno* diventa in tal modo, come è stato osservato, una sorta di romanzo corale, dove il ritmo novellistico del racconto è legato anche alla frantumazione della voce narrante che “coincide con una curiosa voce collettiva, della quale registra umori, giudizi, pettegolezzi, stupori, sarcasmi, cattiverie” [Gioviale, 1984: 162].

### **Disposizione sintagmatica dei tempi e segmentazione del discorso narrativo**

La tipologia dei procedimenti narrativi evidenziati attraverso l'analisi paradigmatica dei tempi verbali, fornisce una chiave per lo studio della strategia compositiva del romanzo messa in atto dall'autore. La distribuzione sintagmatica dei tempi verbali, infatti, permette di individuare come i vari procedimenti si dispongano lungo lo svolgimento sintagmatico del testo in relazione allo sviluppo dell'intreccio.

Al fine di studiare la configurazione sintagmatica dei morfemi temporali, abbiamo assunto come unità testuale minima i capitoli in cui l'autore ha suddiviso il racconto. Si è preferito mantenere questa partizione del livello discorsivo del testo per due motivi: in primo luogo perché ci sembra la più autorevole e significativa; in secondo luogo perché sufficientemente minuziosa da dare un quadro della composizione del testo. Successivamente sono state rilevate le frequenze temporali in ogni capitolo. Gli indici di frequenza si presentano in genere fortemente polarizzati tra due tempi principali. In ogni capitolo il tempo dominante ha una frequenza percentuale oscillante tra il 35 e il 70%, seguito da un tempo secondario, con frequenza tra il 30 e il 20%. Qualora sia presente in misura rilevante un terzo tempo verbale, le frequenze sono sempre inferiori al 20%.

Sulla base delle osservazioni esposte sopra è possibile individuare l'articolazione dei diversi procedimenti narrativi e la loro relazione con le diverse fasi dell'intreccio. Si è potuto così osservare, come ci si poteva aspettare, che ai momenti dell'intreccio che espongono le motivazioni interne e le opinioni dei vari personaggi, si connettono il discorso indiretto libero e le espansioni demarcative diegetiche; le fasi

trasformative vere e proprie invece si distribuiscono tra scene diegetiche, specialmente nella prima parte del romanzo, e scene mimetiche.

Un interesse particolare viene assunto dalle sequenze mimetiche. La distribuzione lessicale dei tempi di primo piano indica come abbiamo rilevato una forte tendenza alla drammatizzazione della narrazione. La collocazione delle scene mimetiche lungo lo svolgimento del testo conferma questa ipotesi. Al fine di evidenziare questa configurazione, si è costruito un prospetto in cui sono elencate sistematicamente le dieci scene mimetiche del romanzo (sia al discorso diretto legato sia a quello libero) con una consistenza lineare rilevante, nell'ordine di successione in cui si trovano nel testo. Nella tabella vengono indicate la collocazione per capitoli, i personaggi implicati e la funzione tematico-narrativa svolta da ogni singola scena dialogata:

Capitolo	Personaggi	Funzione tematico-narrativa
III	Ravi/Pepè	promessa (implicita) di matrimonio
V	Pepè/Ciro	organizzazione del duello
VI	Pepè/Ravi/padrini	organizzazione del duello
XIV	Pepè/Mauro Salvo	minacce di Mauro a Pepè
XVI	Pepè/Ciro	richiesta di aiuto di Pepè a Giro
XVII	Ravi/Stellina/Ciro/Pepè	offerta di aiuto di Giro a Stellina
XVIII	Pepè/Ravi	offerta di aiuto di Giro a Stellina
XIX	Pepè/Ciro	equivoco di Pepè
XXI	Ravi/don Diego	fuga di Stellina
XXIII	Pepè/Ciro	agnizione dell'equivoco

Questo schema mostra come le sequenze sceniche di maggiore estensione si collochino proprio in corrispondenza degli snodi narrativi della *fabula*, i quali vengono dunque “rappresentati” attraverso una prospettiva narrativa con focalizzazione interna ai diversi personaggi implicati nel dialogo. Si deve notare, peraltro, come tutte queste scene (a eccezione di quella nel capitolo XIV), si svolgano in un contesto spaziale riconducibile al *topos* “interno domestico”, proprio come se si trattasse di scene teatrali.

L'andamento complessivo delle frequenze temporali nei capitoli e la successione lineare dei singoli tempi mostra una netta configurazione temporale nella dimensione sintagmatica del romanzo, corrispondente alla segmentazione del testo in sequenze narrative. In primo luogo, si può distinguere molto chiaramente la divisione del testo in due piani enunciativi, uno diegetico, caratterizzato dai tempi commen-

tativi, e l'altro mimetico, caratterizzato dai narrativi, che si articolano paralleli lungo lo svolgimento dell'intreccio.

Poiché nel testo narrativo (malgrado la configurazione enunciativa del *Turno* sia portata al limite) il piano della diegesi pura costituisce il livello testuale dipendente immediatamente dall'atto di enunciazione del narratore, è su questo livello che si proietta la segmentazione della *fabula* in sequenze di eventi con funzione narrativa omogenea. In effetti, nella distribuzione temporale per capitoli si possono riscontrare alcune cesure in corrispondenza delle quali vi sono dei rilevanti cambiamenti negli indici di frequenza dei tempi del gruppo narrativo. In base alla collocazione di queste cesure transizionali, ci sembra di poter dividere il testo in tre macro-sequenze.

I primi tre capitoli, che si caratterizzano per la prevalenza di tempi dello sfondo con forte presenza di presente, costituiscono la prima macro-sequenza narrativa. Essa rappresenta la sequenza di *orientamento* del racconto, che vede infatti la voce narrante alternarsi ai vari personaggi, in special modo il Ravi, per introdurre il tema principale dell'intreccio (la duplice promessa di matrimonio), e le opinioni dei personaggi, compreso il personaggio collettivo "amici e conoscenti", allo stato iniziale del testo. I procedimenti narrativi dominanti sono evidentemente le espansioni demarcative, sia mimetiche sia diegetiche, intervallate al discorso indiretto libero che veicola i soliloqui di don Marcantonio e di Pepè.

I capitoli dal quarto al venticinquesimo, costituiscono la macro-sequenza II, che si può definire la sequenza di *complicazione*. Dal punto di vista temporale si evidenzia una oscillazione che alterna la prevalenza del perfetto a quella del presente, sulla cui base si configura una ulteriore ripartizione in due sotto-sequenze narrative.

La prima si estende lungo i capitoli IV-XIII. Essi sono caratterizzati da una prevalenza del perfetto, mentre l'imperfetto diminuisce; il procedimento narrativo dominante è la scena diegetica (fa eccezione il capitolo V, totalmente mimetico, per il quale valgono le osservazioni fatte poco sopra). Questi capitoli narrano la vicenda dell'amore di Pepè per Stellina, dopo il matrimonio di questa con don Diego. Iniziano infatti con la festa nuziale (cap. IV), e terminano con il vecchio morante, mentre Stellina e Pepè sembrano innamorarsi reciprocamente, intanto che attendono a vegliarlo. Giunti a questo punto dell'intreccio ci si aspetterebbe il lieto fine con una chiusura all'imperfetto. Si apre invece, con il capitolo XIV, una serie di capitoli che costituisce la seconda sequenza di *complicazione*, caratterizzata dalla prevalenza quasi

assoluta del presente, e dunque delle scene mimetiche. A questa sequenza che complessivamente rappresenta un vero e proprio segmento di scrittura drammaturgica, corrisponde dal punto di vista tematico l'intervento di Ciro Coppa, che sconvolge i piani di don Marcantonio e di Pepè, rimettendo in moto l'intreccio.

Questa sequenza si chiude con l'esilarante capitolo XXIV, che vede i tre sconfitti riuniti in casa di don Diego beccarsi l'un l'altro, con una sequenza scenica che evoca un tipico epilogo da commedia delle beffe.

L'ultima sequenza testuale va dal capitolo XXV alla fine, e vede di nuovo la prevalenza dell'imperfetto e dei procedimenti narrativi di sfondo, che si articolano con il presente. Questa sequenza costituisce nel suo complesso la *sequenza risolutiva*. All'imperfetto corrisponde la prevalenza di una configurazione enunciativa con focalizzazione interna ai personaggi, ognuno alle prese con una realtà dimostratasi ingovernabile e imprevedibile: Pepè raccolto nel suo dolore di innamorato abbandonato; il Ravi che ha visto crollare il suo ambizioso sogno di far diventare ricca la figlia; don Diego perseguitato dalla solitudine e dalla sue fobie senili, che vagheggia un ennesimo matrimonio; non sfugge a questa caratterizzazione psicologica nemmeno Ciro Coppa, la cui indole gelosa si trasforma in uno stato paranoico che lo tiene in continua esagitazione.

La conclusione del romanzo, infine, si articola nuovamente sul presente del capitolo XXVIII, che prelude alla rocambolesca chiusura con la morte di Ciro Coppa, e la scena della camera mortuaria, in cui prevale un imperfetto che ricorda l'*imperfetto di rottura* individuato da Weinrich ha nella novellistica in lingua romanza di fine secolo<sup>13</sup>. Ma quasi a ribadire la invadente presenza enunciativa dei personaggi nel testo, l'ultima parola spetta ancora a Marcantonio Ravi, il quale chiude il romanzo ad alta voce, come l'aveva iniziato, testimoniando per l'ennesima volta la "miopia economico-sociale che restringe il suo campo mentale" [Tedesco, 1976: 48].

---

<sup>13</sup> Si veda su questo il capitolo quinto di Weinrich, 1978, dal titolo "I tempi e la messa in rilievo nella novellistica"; in particolare il paragrafo dedicato all'italiano (pp. 153 e segg.) che porta come esempio una novella pirandelliana, *L'uomo solo*.

## Il turno e la crisi del naturalismo

La molteplice scansione del testo narrativo, condotta attraverso la chiave della configurazione temporale, ci permette di affermare che *Il turno* si pone come esperimento letterario di frontiera nella narrativa di Luigi Pirandello. In esso la presenza strutturale della narrativa verista, che rappresenta la spina dorsale delle novelle dei primi anni (si vedano *La ricca*, e le novelle raccolte in *Amori senza amore*)<sup>14</sup>, viene assunta dallo scrittore siciliano come oggetto di una progressiva sperimentazione decostruttiva, che giunge a una completa scomposizione dell'articolazione narrativa del romanzo naturalista.

Questa conclusione è confermata dalla collocazione storico-culturale della genesi del *Turno*, risalente alla metà degli anni novanta. Nel corso dell'ultimo decennio dell'Ottocento comincia a profilarsi, negli ambienti intellettuali italiani, quella "coscienza della crisi", per dirla con Salinari, che caratterizzerà la civiltà europea a cavallo tra i due secoli<sup>15</sup>. Una crisi epocale dell'universo ideologico, culturale ed epistemico che aveva dominato nella seconda metà del secolo scorso. Pirandello si colloca con pieno diritto nell'ambito di questo panorama intellettuale, seppure in una prospettiva fortemente influenzata dalle

---

<sup>14</sup> Pirandello, d'altronde, non abbandona tale architettura nemmeno nel suo capolavoro narrativo, *Il fu Mattia Pascal*, come molti attenti lettori ci hanno attestato, a partire dalla bellissima esegesi dedicata al romanzo da Debenedetti nelle sue lezioni universitarie [Debenedetti, 1971].

<sup>15</sup> Sulla "crisi" di fine secolo, sui rapporti tra questo panorama ideologico-culturale e la produzione letteraria pirandelliana la bibliografia è ormai sterminata, con contributi che vengono dalle più disparate tendenze della critica. Basti ricordare in ambito marxista Salinari [1960], e Leone de Castris [1962 e il saggio in Lauretta (a c. di), 1976], nonché in un lavoro che colloca l'opera pirandelliana entro l'ambito del romanzo europeo Mazzacurati [1987]. Va ricordata naturalmente l'eclettica ermeneutica di Debenedetti, con il suo bellissimo *Il romanzo del novecento* [1971]. Una interessante prospettiva critica, diversa, emerge nella serie di lavori che Renato Barilli, nell'ambito della sua *culturologia*, ha dedicato allo scrittore siciliano, il quale, secondo il critico, non esprime solo e tanto una ideologia della crisi, quanto piuttosto una rifondazione dei paradigmi culturali per l'uomo del novecento, riuscendo a superare in senso positivo il crollo degli orizzonti culturali ottocenteschi: si tratta insomma di un Pirandello fuori della "crisi", che anzi contribuisce alla ricostruzione dell'orizzonte epistemico della modernità. Un lavoro recente che affronta in maniera analitica il rapporto tra la poetica pirandelliana e la "filosofia della crisi" è il saggio di Dombroski in Geerts (a c. di), 1987.

peculiari condizioni sociali e culturali in cui si trova l'Italia post-risorgimentale, e in particolare la sua Sicilia. Ne è testimonianza la sua produzione saggistica coeva<sup>16</sup>.

In questi scritti del decennio '890-'900 la riflessione pirandelliana già dimostra una notevole consapevolezza teorico-critica. Attraverso un attivo confronto con i contemporanei, il giovane Pirandello getta le basi della sua futura produzione artistica e teorica, elaborando una vasta materia, in cui già si possono individuare, sebbene *in fieri*, i nuclei della sua originale poetica, che si svilupperà definitivamente nel primo decennio del '900.

La serie di scritti saggistici, dedicati in varie occasioni da Pirandello ai rappresentanti della letteratura decadentista, dimostra la distanza di poetica e di sensibilità che doveva intercorrere con il simbolismo decadente ed estetizzante di quelli che lo stesso Pirandello definì i "neo-idealisti", in testa tra tutti D'Annunzio. Ma questa distanza dai testimoni nostrani della letteratura decadente ed estetizzante non comporta una parallela, acritica adesione ai canoni del naturalismo. Certamente, l'influsso della cerchia romana dei naturalisti, segnatamente quella di Capuana<sup>17</sup>, ha avuto un peso non indifferente nella maturazione artistica e stilistica del giovane Pirandello<sup>18</sup>. Si può affermare, tuttavia, che, dal punto di vista della poetica, egli prende equamente le distanze tanto dal "neo-idealismo" dell'aulico D'Annunzio, quanto dalla "ingenua o, peggio, artificiosa, dialettalità fonografica dei regionalisti italiani" [Leone de Castris, 1962: 48].

La vicinanza del giovane Pirandello con la cerchia letteraria del Capuana si basa più su un comune atteggiamento di rifiuto polemico

---

<sup>16</sup> Gli articoli citati sono raccolti in *Saggi*. Una puntuale disanima in ordine cronologico della saggistica del giovane Pirandello, che testimonia anche gli influssi di questa sulla prosa narrativa, fino a esempi di intertestualità piuttosto consistenti, si trova Andersson 1966, il quale raccoglie anche molti articoli non figuranti nell'edizione di *Saggi*.

<sup>17</sup> Per una puntuale ricostruzione del rapporto umano e critico-letterario tra il grande narratore e teorico del naturalismo italiano e Pirandello si veda Sipala, 1974. In particolare, relativamente al periodo di cui ci stiamo occupando, e ai rapporti con il circolo di letterati e critici che si raccolse intorno ai due scrittori cfr. pp. 9-11.

<sup>18</sup> E si tratta di un legame decisivo come rileva, tra gli altri, Gioviale: "La ricerca pirandelliana è complicata inizialmente dal debito naturale verso i maestri che gli hanno insegnato a leggere e capire la narrativa, e innanzitutto verso Capuana, che lo ha orientato per tempo nella direzione del romanzo" [Gioviale, 1984: 93-94].

verso la produzione letteraria dannunziana, che su una costruttiva tangenza di poetica e di visione dell'arte. D'altra parte in questo stesso periodo Pirandello ha già maturato, dal punto di vista filosofico, un netto rifiuto dell'ideologia positivista, come più volte egli stesso afferma chiaramente. Basti ricordare, tra gli altri, l'articolo "Arte e coscienza d'oggi" (di cui interi passi sono confluiti nella redazione di *Marta Ayala*<sup>19</sup>), nel quale emerge la precoce coscienza della crisi culturale ed esistenziale che Pirandello avvertiva intorno a sé:

Crollate le vecchie norme, non ancora sorte o bene stabilite le nuove; è naturale che il concetto della relatività d'ogni cosa si sia talmente allargato in noi da farci quasi del tutto perdere l'estimativa. Il campo è libero ad ogni supposizione. L'intelletto ha acquistato una straordinaria mobilità. Nessuno più riesce a stabilirsi un punto di vista fermo e incrollabile. [*Saggi*: 900]

Pirandello riprende questi temi in altri articoli e saggi tra i quali vanno ricordati "Rinunzia", pubblicato nel 1896 su *La critica*, e "Il neo-idealismo" [*Saggi*: 913-921] del 1986, in cui dichiara di condividere con i "simbolisti" le motivazioni profonde della polemica contro il materialismo scienziato del positivismo: "È verissimo che il concetto materialistico del mondo e della vita non appaga più lo spirito moderno, rimasto per esso senza mistero e senza Dio"; e rileva che di fronte al progresso scientifico e tecnologico "la nostra sete rimane tuttavia insaziata, e noi chiediamo ancora: E poi?" [*Saggi*: 916-917].

Questa posizione ideologica e culturale trova riflesso nella speculazione poetica di Pirandello, come abbiamo visto. A questo proposito alcune sintomatiche riflessioni si possono rintracciare in due articoli risalenti al 1895, e dunque coevi al *Turno*, "Come si parla in Italia?" [*Saggi*: 887-891] e "Come si scrive oggi in Italia" [*Saggi*: 891-906], pubblicati sulla rivista di Gino Montaldi *La Critica*. In entrambi l'Autore, riprendendo peraltro temi già affrontati in scritti anteriori, lamenta la mancanza di "spontaneità" e "sincerità" nella prosa e nella lingua letteraria italiana in genere. Si è detto sintomatiche poiché la rivendicazione di "sincerità" e di "spontaneità" nel fare artistico, come ha evidenziato Jurii Lotman, caratterizza le poetiche di transizione che

---

<sup>19</sup> Si veda per una puntuale disanima di questi passaggi intertestuali ancora Andersson 1966.

si sviluppano nelle fasi di crisi dei sistemi culturali, definite dal grande studioso russo “estetiche della contrapposizione”:

Ai metodi di simulazione della realtà ai quali il lettore è abituato, l'artista contrappone la propria soluzione originale, che considera più vera... Soggettivamente la tendenza dell'autore e del suo pubblico a distruggere il sistema di regole consuete può essere inteso come il rifiuto di ogni norma strutturale, come l'operare «senza regole». [Lotman, 1990: 341-342]

Caratteristica delle epoche transizione tra paradigmi culturali, è appunto la destrutturazione dei codici culturali dominanti, che si riflette all'interno dei sistemi modellizzanti secondari, come quello letterario, nell'abbattimento dei confini tra i generi, nella commistione tra stili e strutture, spinta dall'esigenza di ricercare un nuovo linguaggio.

Questa destrutturazione è realizzata da Pirandello per vie interne: ora infrangendo i canoni tematici, ora le strutture formali e compositive dell'intreccio, ora l'equilibrio della distanza narrativa, ora, infine, rifiutando l'assenza dell'autore dalla narrazione (carattere imprescindibile della poetica naturalista). E l'immanenza dell'autore nel meccanismo narrativo viene messa in atto da Pirandello sia dove assume come controfigura un io-narrante pervasivo – attraverso l'uso della prima persona che con il suo pressante intervento raziocinante scompiglia la sequenzialità dell'intreccio<sup>20</sup> – sia dove si immerge nella corralità macchietistica di un chiacchericcio provinciale, come avviene nel *Turno*.

L'incessante cambiamento della configurazione enunciativa di questo romanzo, dunque, più che una innata attitudine verso la scrittura drammaturgica, segnala una tendenza alla commistione dei generi<sup>21</sup>. L'espansione della dialogicità, fino ai limiti della teatralizzazione, costituisce lo strumento attraverso il quale Pirandello opera un primo tentativo di superamento della struttura narrativa naturalista, come ha rilevato anche Guglielminetti:

---

<sup>20</sup> Riguardo alle caratteristiche formali di questo verboso narratore intradiegetico, che risulta ben lontano dall'onnisciente narratore manzoniano, progressivamente trasformandosi in “personaggio tra gli altri personaggi” della narrazione, si veda l'interessante articolo di C. Van Der Voort “Tecnica narrativa e poetica dell'umorismo” in Geerts (a c. di), 1987.

<sup>21</sup> A questo proposito si veda anche Gioviale, 1984: 111.

È probabile... che col *Turno* egli abbia voluto dimostrare polemicamente a critici e lettori il grado di dissolvimento a cui poteva essere condotta l'impalcatura del romanzo naturalistico, una volta soltanto che si fosse accentuata la prevalenza delle parti dialogate a scapito di quelle descrittive. [Guglielminetti, 1986: 65]

La voce narrante, immergendosi in quella coralità discorsiva più volte evidenziata nel corso dell'analisi, produce un gioco disgregante dei punti di vista dei personaggi sulla realtà. Questa non viene presentata come oggetto dell'atto di enunciazione del narratore che, pur nella sua impersonalità, garantisca sulla sua consistenza oggettiva, come avveniva nella narrativa naturalista. Piuttosto, si assiste a un continuo succedersi di eventi e azioni mediati attraverso la prospettiva dei vari personaggi, ognuno più o meno impegnato a far prevalere il suo punto di vista<sup>22</sup>. La struttura enunciativa tipicamente naturalista subisce un effetto di straniamento, attraverso il quale si attua la negazione dell'assioma stesso del naturalismo, quello dell'oggettività del reale.

Dal medesimo intento destrutturante deriva, parallelamente, la progressiva rifunzionalizzazione del sistema dei tempi verbali che abbiamo evidenziato nel corso della nostra analisi. Questo processo agisce sul sistema temporale del romanzo, attraverso un cambiamento di segno nella funzione dei tempi verbali. Già Terracini [1966] nota l'uso particolare del "presente di sogno" nel discorso indiretto libero; e ancora Guglielminetti [1986] individua come il condizionale diventi "il tempo del soliloquio". Più recentemente, Claude Cazalé, lavorando sul corpus novellistico, ha riassunto in questo modo il complessivo straniamento della sintassi narrativa pirandelliana:

Il sistema tradizionale dei tempi si complica: all'alternarsi per la reciproca messa in rilievo del Passato remoto (azioni-cardine decisive per lo sviluppo del processo, puntuali e trasformative) e degli Imperfetti (azioni durative e ripetitive) e Piuच्chepperfetti (antecedenti, *flashes-back*) si aggiungono un ricorso ai tempi del passato o al condizionale per il discorso indiretto libero e l'uso di forme del presente nel soliloquio. [Cazalé, 1989: 85]

---

<sup>22</sup> In questo senso si ritrovano i segni di quel carattere retorico della narrativa pirandelliana di cui ha parlato Renato Barilli, in ragione della quale "gli elementi riguardanti la presentazione di fatti e personaggi, le battute di dialogo, i monologhi interiori devono assumere anch'essi un tono il più possibile persuasivo e accomodante" [Barilli, 1988: 137]

Va situato in questa dinamica strutturale anche il carattere eminentemente comico del *Turno*, carattere che è stato messo in rilievo da molta la critica e sul quale non ci soffermiamo. Lo stesso Pirandello, nella *Premessa* alla riedizione del 1915 presso Treves, definì le vicende narrate nel racconto come “gaje” e “liete”, determinazioni che evocano un atteggiamento del narrare di natura comica, assai distanti dalla dominante tragica che informava la narrativa verista, in special modo quella del Capuana, ma altrettanto difficilmente riconducibili al posteriore *umorismo* pirandelliano, nel quale il momento del tragico è tanto essenziale quanto quello del riso. In questo senso bene definisce la natura e anche l’anomalia del romanzo nel panorama della narrativa pirandelliana, Renato Barilli quando scrive:

... è questa una prova generale, e delle più intense e riuscite, di tutti i meccanismi “comici” che poi l’Autore spargerà a piene mani in decine di novelle e drammi. Non è raggiunta invece la soglia dell’umorismo, proprio perché nessuno dei personaggi di questa vicenda ragiona sui casi eccezionali che stanno accadendo, ne li fa risultare da una consapevole rivolta contro le norme. [Barilli, 1986:150]

La comicità, conseguita attraverso una serie di procedimenti strutturali sia al livello di configurazione sia a livello di intreccio e di semantica del carattere, diventa strumento di dissacrazione dei riferimenti assiologici e ideologici che orientavano il sistema culturale in via di disgregazione.

Nella successione diacronica delle strutture letterarie, uno dei generi che presenta i caratteri della comicità e della ambivalenza nella struttura di genere, è la *parodia*, in cui “l’autore, con determinati elementi di costruzione della sua opera, suscita nel lettore quella struttura che sarà poi annullata” [Lotman, 1990: 343]. *Il turno*, con la sua dialettica strutturale tra procedimenti naturalistici e procedimenti anti-naturalistici, con la farsesca dissacrazione delle convenzioni sociali della borghesia provinciale, con la sua comicità paradossale, rientra pienamente nella definizione di *parodia del naturalismo*. Parodia in cui è possibile rintracciare i momenti iniziali di un processo che, portato a termine nell’ultimo romanzo, *Uno, nessuno e centomila*, sancirà, oltre trent’anni più tardi, non solo l’esito del realismo, ma l’impossibilità stessa di qualsiasi narrazione.



## Bibliografia

- Andersson G., 1966, *Arte e teoria. Studi sulla poetica del giovane Pirandello*, Almqvist & Wiksell, Uppsala.
- Angelini F., 1976, «Luigi Pirandello», in Angelini et al., *La letteratura italiana. Storia e testi*, vol. IX, tomo I, *Il Novecento. Dal decadentismo alla crisi dei modelli*, Laterza, Roma-Bari.
- Ausiello G. (et al.), 1991, *Modelli e linguaggi dell'informatica*, McGraw-Hill Italia, Milano.
- Barilli R., 1986, *Pirandello. Una rivoluzione culturale*, Mursia, Milano.
- Barilli R., 1988, *La linea Svevo-Pirandello*, Mursia, Milano.
- Barnard D.T., Fraser C.A., Logan G.M., 1988, *Generalized Markup for Literary Texts*, in «Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing», vol. 3, pp. 26-31.
- Barnard, D.T. (et al.), 1988, «SGML-Based Markup for Literary Texts: Two Problems and Some Solutions», *Computers and the Humanities*, 22, pp. 265-276.
- Berners-Lee T., 2001, *L'architettura del nuovo Web*, Feltrinelli, Milano.
- Bertinetto P.M., 1986, *Tempo, aspetto e azione nel verbo italiano*, Accademia della Crusca, Firenze.
- Bolter, J. D., 1991, *Writing Space: The Computer, Hypertext, and the History of Writing*, Lawrence Erlbaum Associates.
- Bryan, M., 1988, *SGML: An Author's Guide to the Standard Generalized Markup Language*, Addison-Wesley, Wokingham/Reading/New York.
- Burnard L., 1995a, *Text Encoding for Information Interchange. An Introduction to the Text Encoding Initiative*, Oxford University Computing Services, Oxford, <<http://www.tei-c.org/Papers/J31>>.
- Burnard L., 1995b, *What is SGML and How Does It Help?*, in Ide, Véronis (a cura di) 1995, pp. 41-50.
- Burnard L., Greenstein D., 1995, *Speaking with one voice: Encoding Standards and the Prospects for an Integrated Approach to Computing in History*, in Ide, Véronis, 1995, pp. 137-48.
- Burnard L., Sperberg-McQueen C.M. (a c. di) 2002, *Guidelines for Electronic Text Encoding and Interchange*, TEI Consortium, Oxford-Providence-Charlottesville-Bergen, <<http://www.tei-c.org/P4X/>>.
- Buzzetti D., 1995, *Image Processing and the Study of Manuscript Textual Traditions*, «Historical Methods», 28 (1995), pp. 145-163.
- Buzzetti D., 1996, *Il testo 'fluidò': Sull'uso dell'informatica nella critica e nell'analisi testuale*, in Floridi (a c. di), 1996, pp. 85-93.

- Buzzetti D., 1999, *Rappresentazione digitale e modello del testo*, in *Il ruolo del modello nella scienza e nel sapere*, Atti del Convegno (Roma, 27-28 ottobre 1998), Accademia Nazionale dei Lincei, Roma, pp. 127-61.
- Buzzetti D., 2002, *Digital Representation and the Text Model*, in «New Literary History», vol. 33, n. 1, pp. 61-68.
- Calvo M., Ciotti F., Roncaglia G., Zela M.A., 2003, *Internet 2004. Manuale per l'uso della rete*, Laterza, Roma–Bari.
- Caton P., 2005, *LMNL Matters?*, in *Proceedings of Extreme Markup Languages 2005*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2005/Caton01/EML2005Caton01.html>>.
- Cazalé C. (a c. di), 1993, “Narrativa”, n.4 (marzo 1993), *Fine della storia e storie senza fine*.
- Cazalé C., 1989, *Tempo, Azione, Identità: costanti narrative nella raccolta Scialle Nero*, in “Rivista di studi pirandelliani”, 2 (III S).
- Cazalé C., Mordenti R., 1997, *La costituzione del testo e la «comunità degli interpreti»*, in Nerosi Bellman (a c. di), 1997, pp. 13-38.
- Ciotti F. (a c. di), 2005, *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari*, Bonnard, Milano.
- Ciotti F., 1994, *Il testo elettronico: memorizzazione, codifica e edizione informatica del testo*, in Leonardi, Morelli, Santi (a c. di), 1994.
- Ciotti F., 1995, *Testi elettronici e banche dati testuali. Problemi teorici e tecnologie*, in «Schede umanistiche», N.S., 2, pp. 147-78.
- Ciotti F., 1997a, *Cosa è la codifica informatica dei testi?*, in Gruber, Paoletto (a c. di), *Umanesimo & Informatica*, Metauro, Pesaro, 1997, pp. 55-85.
- Ciotti F., 1997b, *Testo rappresentazione e computer. Contributi per una teoria della codifica testuale*, in Nerosi Bellman (a c. di), 1997, pp. 213-230.
- Ciotti F., 2001, *Text encoding as a theoretic language for literary text analysis*, in Fiormonte, Usher (a c. di) 2001, pp 39-47.
- Ciotti F., 2002, *Tempi verbali e struttura narrativa*, in Mordenti (a c. di), 2002.
- Ciotti F., 2003, *Teoria, progetto e implementazione di una biblioteca digitale. Testi Italiani on-Line*, in Fiormonte (a c. di), 2003.
- Ciotti F., Fiormonte D., 2002, *Informatica e letteratura*, in Barberi Squarotti G. (et. al.), *Letteratura. Progetto modulare, Strumenti di analisi e scrittura*, Tomo 0, Atlas, Bergamo, pp. 153-170.
- Ciotti F., Roncaglia G., 2000, *Il mondo digitale. Introduzione ai nuovi media*, Laterza, Roma–Bari.
- Conte M.E. (a c. di), 1988, *La linguistica testuale*, Feltrinelli, Milano.
- Coombs J.H., De Rose S.J., Renear A. 1987, *Markup and the Future of the Scholarly Text Processing*, in «Communications of the Association for Computing Machinery», vol. 30, n. 11, pp. 933-47, poi in Landow,

- Delany (a c. di), 1993, pp. 85-118, <<http://www.oasis-open.org/cover/coombs.html>>.
- Corsinovi G., 1979, *Pirandello e l'Espressionismo. Analogie culturali e anticipazioni espressive nella prima narrativa*, Tilgher, Genova.
- Corsinovi G., 1983, *Pirandello: tradizione e trasgressione. Studi su Pirandello e la letteratura italiana tra '800 e '900*, Tilgher, Genova.
- Coveri L. (a c. di), 1984, *Linguistica testuale*, Bulzoni, Roma.
- De Mauro, T., 1992, *Informatica e linguistica*, in *Calcolatori e scienze umane*, Milano, Etas, pp. 195-210.
- De Rose S.J. (et al.), 1990, *What is Text, Really?*, in «Journal of Computing in Higher Education», vol. 1, n. 2, pp. 3-26.
- Debenedetti G., 1945, *Una giornata di Pirandello*, in idem, *Saggi Critici. Serie Nuova*, Mondadori, Milano 1945.
- Debenedetti G., 1971, *Il romanzo del novecento*, Garzanti, Milano.
- Dennet, D., 1995, *The Interpretation of Texts, People and Other Artifacts, Philosophy and Phenomenological Research*, L, Supplement, Fall 1990, pp. 177-194, poi in M. Losonsky, ed., *Language and Mind: Contemporary Readings in Philosophy and Cognitive Science*, Blackwells, 1995, <<http://www.tufts.edu/as/cogstud/papers/intrptxt.htm>>.
- DeRose S., 2004, *Markup Overlap: A Review and a Horse*, in *Proceedings of Extreme Markup Languages 2004*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2004/DeRose01/EML2004DeRose01.html>>.
- DeRose, S. J., Durand, D., Mylonas, E., Renear, A. H., 1990, «What is Text, Really?», *Journal of Computing in Higher Education*, 1/2 (1990), pp. 3-26.
- Di Pietro A., 1979, *Luigi Pirandello e la «crisi della ragione»*, in *Novecento. I contemporanei*, vol. III, Marzorati, Milano 1979, pp. 2195-2234.
- DuCharme B., 1999, *XML: the annotated specification*, Prentice Hall, Upper Saddle River.
- Durusau P., O'Donnell, M. B., 2004, *Tabling the Overlap Discussion*, in *Proceedings of Extreme Markup Languages 2004*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2004/Durusau01/EML2004Durusau01.html>>.
- Eco U., 1975, *Trattato di semiotica generale*, Bompiani, Milano, (X ed., 1987).
- Eco U., 1979, *Lector in fabula*, Bompiani, Milano.
- Eco U., 1990, *I limiti dell'interpretazione*, Bompiani, Milano.
- Escarpit, R., 1976, *Théorie générale de l'information et de la communication*, Paris, Librairie Hachette, 1976 (trad. it. di Maria Grazia Rombi, *Teoria dell'informazione*, Editori Riuniti, Roma, 1979).
- Finneran R., 1996, *The Literary Text in the Digital Age*, University of Michigan Press, Ann Arbor.

- Fiormonte D., 2003, *Scrittura e filologia nell'era digitale*, Bollati Boringhieri, Torino.
- Fiormonte D. (a c. di), 2003, *Informatica umanistica: dalla ricerca all'insegnamento*, Bulzoni, Roma.
- Fiormonte D., Usher J. (a c. di), 2001, *New Media and the Humanities: Research and Applications*, Oxford University Humanities Computing Unit, Oxford.
- Floridi L. (a c. di), 1996, *Filosofia & informatica*, Torino, Paravia.
- Fondazione IBM., 1992, *Calcolatori e scienze umane*, Scritti del Convegno organizzato dall'Accademia Nazionale dei Lincei e dalla Fondazione IBM Italia, Etas Libri, Milano.
- Gardin J.C., 1991, *Le calcul et la raison. Essai sur la formalisation du discours savant*, Édition de l'École des Hautes Études en Sciences Sociales, Paris.
- Geerts W., Musarra F. (a c. di), 1987, *Luigi Pirandello: poetica e presenza*, Leuven University Press-Bulzoni, Leuven-Roma.
- Genette G., 1976, *Figure III. Discorso del racconto*, Torino, Einaudi.
- Genette G., 1987, *Soglie. I dintorni del testo*, Torino, Einaudi.
- Gigliozzi G., 1987a, «Codice, testo e interpretazione», Gigliozzi, G. (a c. di), 1987, pp. 65-84.
- Gigliozzi G., 1991 *Pirandello, la novella e il gioco del rovescio*, in «Rivista di studi pirandelliani», 6/7 (III S), pp. 73-85.
- Gigliozzi G., 1992, «Modellizzazione delle strutture narrative», in Fondazione IBM, 1992, pp. 303-314.
- Gigliozzi G., 1997, *Testo e computer*, Bruno Mondadori, Milano.
- Gigliozzi G., 2003, *Introduzione all'uso del computer negli studi letterari*, a c. di F. Ciotti, Bruno Mondadori, Milano.
- Gigliozzi G. (a c. di), 1987, *Studi di codifica e trattamento automatico di testi*, Bulzoni, Roma.
- Gigliozzi G., Giuliani S., 1993a *Una parola che non dice nulla*, in Cazalé 1993, pp. 87-89.
- Gigliozzi G., Giuliani S., 1993b *La rete delle formiche. Un applicazione di SEBNET*, in Cazalé 1993, pp. 137-157.
- Gigliozzi G., Giuliani, S., Sensini P., 1987 *SEB – Sistema esperto per l'analisi di brani. Per un'analisi automatica di fiabe*”, in Gigliozzi (a c. di) 1987.
- Gioviale F., 1984, *La poetica narrativa di Pirandello. Tipologia e aspetti del romanzo*, Patron, Bologna.
- Giuliani, S., Sensini P., 2002, *La narrativa dentro il calcolatore*, in Mordenti (a c. di), 2002, pp.110-122.
- Goldfarb C.F., 1981, «A Generalized Approach to Document Markup», in *Proceedings of the ACM SIGPLAN—SIGOA Symposium on Text Manipulation*, ACM, New York.

- Goldfarb C.F., 1990, *The SGML Handbook*, Oxford University Press, Oxford.
- Goldfarb C.F., Prescod P., 2002, *XML handbook*, Prentice Hall, Upper Saddle River.
- Goodman N., 1968 (2003), *I linguaggi dell'arte*, Net, Milano (ed. precedenti il Saggiatore, 1976, 1998).
- Greimas A., Courtés J., 1979, *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Hachette, Paris.
- Guglielminetti M., 1986, *Il romanzo del novecento italiano. Struttura e sintassi*, Editori Riuniti, Roma.
- Gulbransen D., 2002, *XML schema*, McGraw-Hill, Milano.
- Herwijnen E., 1994, *Practical SGML*, Kluwer Academic Publishers, Dordrecht.
- Hockey S., 2000, *Electronic Texts in the Humanities*, Oxford University Press, Oxford-NewYork.
- Huitfeldt C., 1995, *Multi-Dimensional Texts in a One-Dimensional Medium*, in «Computer and the Humanities», vol. 28, pp. 235-41.
- Huitfeldt C., Sperberg-McQueen C.M., 2004, *GODDAG: A Data Structure for Overlapping Hierarchies*, in «Lecture Notes in Computer Science, vol. 2003/2004», pp. 139-160.
- Huitfeldt C., Sperberg-McQueen C.M., 2006, *Representation and processing of Goddag structures: implementation strategies and progress report*, in *Proceedings of Extreme Markup Languages 2006*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2006/Huitfeldt01/EML2006Huitfeldt01.html>>
- Ide N., Greenstein D. (a c. di), 1999, «Computers and the Humanities» (Special double issue. Tenth anniversary of the Text Encoding Initiative), vol. 33, nn. 1-2.
- Ide N., Hockey S. (a c. di), 1996, *Research in Humanities Computing*, Oxford University Press, Oxford.
- Ide N., Véronis J. (a c. di), 1995, *The Text Encoding Initiative: Background and Context*, Kluwer Academic Publishers, Dordrecht.
- Kay M., 2004, *XSLT 2.0 Programmer's Reference*, Wrox Press/Wiley, Indianapolis (IN).
- Lancashire I., 1996, *Using Tact with Electronic Texts*, MLA, New York.
- Landow G.P., Delany, P. (a c. di.), 1991, *Hypermedia and Literary Studies*, The MIT Press, London.
- Landow G.P. (a c. di), 1994, *Hyper/text/theory*, Johns Hopkins University Press, Baltimore.
- Landow G.P., 1998, *Iper testo. Tecnologie digitali e critica letteraria*, Bruno Mondadori, Milano.
- Landow G.P., Delany P. (a c. di), 1993, *The Digital Word: Text-Based Computing in the Humanities*, MIT Press, Cambridge (MA).
- Lauretta E. (a c. di), 1976, *Il romanzo di Pirandello*, Palumbo, Palermo.

- Lavagnino J., 1996, *Completeness and Adequacy in Text Encoding*, in Finneran 1996, pp. 63-76.
- Lavinio C., 1984, *L'uso dei tempi nelle fiabe orali e scritte*, in Coveri 1984, pp. 289-306.
- Leonardi C., Morelli M., Santi F. (a c. di), 1994, *Macchine per leggere. Tradizioni e nuove tecnologie per comprendere i testi*, Fondazione Enzo Franceschini – Centro Studi sull'Alto Medioevo, Spoleto–Todi.
- Leone de Castris A., 1962, *Storia di Pirandello*, Laterza, Roma–Bari.
- Lo Cascio V., 1987, *Strutture temporali e strategie poetiche in Pirandello*, in Geerts, Musarra (a c. di) 1987, pp. 159-190.
- Lotman J., 1990, *La struttura del testo poetico*, Mursia, Milano.
- Lotman J., Uspenskij B.A., 1975, *Tipologia della cultura*, Bompiani, Milano.
- Mazzacurati G., 1987, *Pirandello nel romanzo europeo*, Il Mulino, Bologna.
- McCarty W., 2005, *Humanities Computing*, Palgrave MacMillan, London.
- McGann J., 2001, *Radiant Textuality. Literature after the World Wide Web*, Palgrave, New York.
- McGann J., 2004, *What is Already Encoded by the Text*, in Unsworth J., Schreibman S., Siemans R. (a c. di), 2004.
- McLuhan M., 1962, *The Gutenberg Galaxy: The Making of Typographic Man*, University of Toronto Press, Toronto.
- Megginson D., 1998, *Structuring XML Documents*, Prentice Hall, Upper Saddle River (NJ).
- Minsky M.L., 1968, *Matters, Mind and Models*, in Minsky, (ed.), *Semantic Information Processing*, Cambridge, M.I.T. Press, 1968 (tr. it., "Materia, mente e modelli", in Somenzi & Cordeschi, (eds.), 1986, pp. 307-315).
- Mordenti R. (a c. di), 2002, *Giuseppe Gigliozzi: la fondazione dell'informatica applicata al testo letterario*, numero monografico di «Testo & Senso», nn. 4-5.
- Mordenti R., 2002, *Informatica e critica dei testi*, Bulzoni, Roma.
- Nerozzi Bellman P. (a c. di), 1997, *Internet e le Muse*, Mimesis, Milano.
- Numerico T., Vespignani A., 2003, *Informatica per le scienze umanistiche*, Il Mulino, Bologna.
- Orlandi T. (a c. di), 1993, *Discipline umanistiche e informatica. Il problema dell'integrazione*, Accademia Nazionale dei Lincei, Roma.
- Orlandi T. (a c. di), 1997, *Discipline umanistiche e informatica. Il problema della formalizzazione*, Roma, Accademia Nazionale dei Lincei.
- Orlandi T., 1990, *Informatica umanistica*, La Nuova Italia scientifica, Roma.
- Orlandi T., 1999, *Ripartiamo dai diasistemi*, in *I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici*, Accademia Nazionale dei Lincei, Roma, pp. 87-101.
- Orlandi T., 2002, *Is Humanities Computing a Discipline?*, in «Jahrbuch für Computerphilologie», vol. 4, pp. 51-58, <<http://computerphilologie.uni-muenchen.de/jg02/orlandi.html>>.

- Petőfi J.S., 1974, *Semantic, Pragmatics, Text Theory*, Università di Urbino, Centro Internazionale di semiotica e linguistica, Working papers, Serie A, 36, 1974 (tr. it. *Semantica, pragmatica, teoria del testo*, in Conte 1977, pp. 195-223)
- Pichler A., 1995a, «Advantages of a Machine-Readable Version of Wittgenstein's Nachlass», Johannessen and Nordenstam (eds.) *Culture and Value: Philosophy and the Cultural Sciences*, The Austrian Ludwig Wittgenstein Society (1995)
- Pichler A., 1995b «Transcriptions, Texts and Interpretations», Johannessen and Nordenstam (eds.) *Culture and Value: Philosophy and the Cultural Sciences*, The Austrian Ludwig Wittgenstein Society (1995)
- Pirandello L., 1960, *Saggi, poesie, scritti vari*, (a c. di Manlio Lo Vecchio Musti), Mondadori, Milano.
- Pirandello L., 1990, *Tutti i romanzi*, (a c. di Giovanni Macchia), "I Meridiani", Mondadori, Milano.
- Putnam H., 1993, *Rappresentazione e realtà*, Garzanti, Milano.
- Putnam H., 1994, *Ragione verità e storia*, Il Saggiatore, Milano.
- Rahitz S., Walsh N., Burnard L., 2004, *A unified model for text markup: TEI, Docbook, and beyond*, XML Europe 2004 – Conference Proceedings, IDEAlliance, <[http://www.idealliance.org/papers/dx\\_xml04/papers/03-08-01/03-08-01.pdf](http://www.idealliance.org/papers/dx_xml04/papers/03-08-01/03-08-01.pdf)>.
- Ravazzoli, F., 1984, *L'altro figlio di Pirandello dalla novella all'atto unico: un'occasione tipologica*, in Coveri 1984, pp. 315-340.
- Renear A., 1997, *Out of Praxis: Three (Meta)Theories of Textuality*, in Sutherland 1997, pp. 107-26.
- Renear A., 2000, *The Descriptive/Procedural Distinction is Flawed*, in «Markup Languages: Theory and Practice», vol. 2, n. 4, pp. 411-20.
- Renear A., 2004, *Text Encoding*, in Unsworth J., Schreibman S., Siemens R. (a c. di), 2004.
- Renear A., Durand D., Mylonas E., 1995, *Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies*, in Ide, Hockey (a c. di) 1996; versione preliminare 1992, <<http://www.stg.brown.edu/resources/stg/monographs/ohco.html>>.
- Robinson P.M.W., 1994, *The Transcription of Primary Textual Sources Using SGML*, Office for Humanities Communication, Oxford.
- Robinson P.M.W., 1997, *New Directions in Critical Editing*, in Sutherland 1997, pp. 145-71.
- Roncallo A., 1984, *Variabili nella tipologia dei testi letterari: romanzo e racconto*, in Coveri (a c. di) 1984, pp. 153-156.
- Rosenblueth A., Wiener N., 1942, *The Role of Models in Science*, in «Phil.Sci.», Vol. 12, pp. 316-21 (tr. it. in Somenzi, Cordeschi 1994, pp. 86-92)
- Rusty Harold E., Scott Means W., 2001, *XML. Guida di riferimento*, Apogeo-O'Reilly, Milano.

- Salinari C., 1960, *La coscienza della crisi*, in idem, *Miti e coscienza del decadentismo europeo*, Feltrinelli, Milano.
- Schonefeld O., Witt A., 2006, *Towards validation of concurrent markup*, in *Proceedings of Extreme Markup Languages 2006*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2006/Schonefeld01/EML2006Schonefeld01.html>>
- Segre C., 1974, *Le strutture e il tempo*, Einaudi, Torino.
- Segre C., 1979, *Semiotica filologica. Testo e modelli culturali*, Einaudi, Torino.
- Segre C., 1981, *Punto di vista e polifonia nell'analisi narratologica*, in *Atti del Convegno internazionale «Letterature classiche e narratologia»*, Istituto di filologia latina dell'Università di Perugia, pp. 51-65 (ora in Segre, 1984, pp. 85-101).
- Segre C., 1984, *Teatro e romanzo*, Einaudi, Torino.
- Segre C., 1985, *Avviamento all'analisi del testo letterario*, Einaudi, Torino.
- Sipala P.M., 1974, *Capuana e Pirandello. Storia e testi di una relazione letteraria*, Bonanno, Catania.
- Somenzi V., Cordeschi R. (a c. di), 1994, *La filosofia degli automi. Origini dell'intelligenza artificiale*, Bollati-Boringhieri, Torino (pr. ed. 1986).
- Sperberg-McQueen C.M., 1991, *Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts*, in «Literary and Linguistic Computing», vol. 6, n. 1, pp. 34-46.
- Sperberg-McQueen C.M., 1994, *Textual Criticism and the Text Encoding Initiative*, relazione presentata alla conferenza MLA 1994 (San Diego, Dicembre 1994), <<http://www.tei-c.org/Vault/XX/mla94.html>>; poi in Finneran 1996, pp. 37-61.
- Sperberg-McQueen C.M., Burnard L., 1995, *The Design of the TEI Encoding Scheme*, in Ide, Veronis 1995, pp. 17-40.
- Sperberg-McQueen C.M., Huitfeldt C., Renear A., 2000, *Meaning and Interpretation of Markup*, in «Markup Languages: Theory and Practice», vol. 2, n. 3, pp. 215-34, <<http://www.w3.org/People/cmsmcq/2000/mim.html>>.
- Sutherland K. (a c. di), 1997, *Electronic Text. Investigations in Method and Theory*, Clarendon Press, Oxford.
- Tedesco N., 1976, *Memoria genetica ed entropia farsesco-espressionistica nel Turno di Pirandello*, in Lauretta (a c. di) 1976, pp. 37-47.
- Tennison J., Piez W., 2002, *The Layered Markup and Annotation Language (LMNL)*, in *Proceedings of Extreme Markup Languages 2002*, <<http://www.idealliance.org/papers/extreme/proceedings/html/2002/Tennison02/EML2002Tennison02.html>>.
- Terracini B., 1966, *Le Novelle per un anno di Luigi Pirandello*, in idem, *Analisi stilistica. Teoria, storia, problemi*, Feltrinelli, Milano 1966, pp. 283-385.

- Thaller, M., 1992, *Images and Manuscripts in Historical Computing*, Max Planck-Institute für Geschichte, Scripta Mercature Verlag, St. Katharinen.
- Thaller, M., 1993, *κλεισ. A Database System*, Max Planck-Institute für Geschichte, Scripta Mercature Verlag, St. Katharinen.
- Unicode Consortium 2003, *The Unicode Standard, Version 4.0*, Addison-Wesley, Reading (MA).
- Unsworth J., Schreibman S., Siemans R. (a cura di), *Blackwell's Companion to the Digital Humanities*, Blackwell, Oxford.
- Van der Vlist, E. 2003, *Relax NG*, O'Reilly, Sebastopol (CA), <<http://books.xmlschemata.org/relaxng/>>.
- W3C Consortium 2004, *Extensible Markup Language (XML) 1.0*, a cura di Bray T., Paoli J., Sperberg-McQueen C.M., Maler E., W3C Consortium, <<http://www.w3.org/TR/REC-xml/>>.
- Weinrich H., 1974, *Textsyntax des französischen Artikels*, in Kallmeyer, W. et al., (a c. di), *Lektürekolleg zur Textlinguistik*, Band II, Athenäum Fischer Taschenbuch Verlag, Frankfurt am Main, pp. 267-284 (trad. it. *Sintassi testuale dell'articolo francese*, in Conte 1988, pp. 53-65).
- Weinrich H., 1978, *Tempus. Le funzioni dei tempi nel testo*, Il Mulino, Bologna.
- Weinrich H., 1988 *Lingua e linguaggio nei testi*, Feltrinelli, Milano.



AREE SCIENTIFICO-DISCIPLINARI

Area 01 – Scienze matematiche e informatiche

Area 02 – Scienze fisiche

Area 03 – Scienze chimiche

Area 04 – Scienze della terra

Area 05 – Scienze biologiche

Area 06 – Scienze mediche

Area 07 – Scienze agrarie e veterinarie

Area 08 – Ingegneria civile e Architettura

Area 09 – Ingegneria industriale e dell'informazione

**Area 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche**

Area 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche

Area 12 – Scienze giuridiche

Area 13 – Scienze economiche e statistiche

Area 14 – Scienze politiche e sociali

*Le pubblicazioni di Aracne editrice sono su*

[www.aracneeditrice.it](http://www.aracneeditrice.it)



Finito di stampare nel mese di aprile del 2007  
dalla tipografia «Braille Gamma S.r.l.» di Santa Rufina di Cittaducale (Ri)  
per conto della «Aracne editrice S.r.l.» di Roma

CARTE

Copertina: *Digit Linen* 270 g/mq  
Interno: *Usomano bianco Selena* 80 g/mq  
Allestimento: legatura a filo di refe / brossura

