

ACCADEMIA NAZIONALE DEI LINCEI

ANNO CCCXCVI - 1999

CONTRIBUTI DEL
CENTRO LINCEO INTERDISCIPLINARE
«BENIAMINO SEGRE»
N. 100

IL RUOLO DEL MODELLO
NELLA SCIENZA E NEL SAPERE

(Roma, 27-28 ottobre 1998)



ROMA

ACCADEMIA NAZIONALE DEI LINCEI

1999

DINO BUZZETTI *

RAPPRESENTAZIONE DIGITALE E MODELLO DEL TESTO

ABSTRACT. — *Text Modelling and Digital Text Representation.* — The adequacy of a digital text representation depends both on the exhaustive character of the kind of markup which renders textual structure explicit and its analytical functionality. The encoding practice conforming to the *Guidelines* of the TEI is affected by the limitations of the form of strongly embedded markup inherited from its adopted markup language (SGML): the structure assignable to the text ends up depending on the linear structure of its *expression* and cannot fully represent the form of its *content*, generally non-linear. At any rate, markup does not define a data model; the operational model necessary to formalizing analytical procedures has to be defined independently from markup and is best represented through kinds of weakly embedded markup. Understanding the logical status of markup, however, is essential to its critical employment. Markup makes explicit self-referring features of the text, where the mobile character of the relationship between expression and content shows itself: markup makes it possible to represent, in a text, the dynamic compensation between the structure of the expression (textual variants) and the structure of the content (text interpretations).

RIASSUNTO. — L'adeguatezza della rappresentazione digitale del testo dipende dall'esaustività e dalla funzionalità all'analisi che possono essere ottenute con la forma di codifica (*markup*) che ne esplicita la struttura. La pratica di codifica ispirata alle *Guidelines* della TEI risente delle limitazioni della forma di *markup* vincolato del linguaggio di codifica adottato (SGML). La struttura assegnabile al testo è vincolata alla struttura lineare dell'espressione e non può esprimere, in generale, la forma non lineare del contenuto. Il *markup* non definisce un modello di dati: il modello necessario alla formalizzazione delle operazioni analitiche deve essere definito indipendentemente dal *markup* ed è più adeguata-

* Dipartimento di Filosofia - Università degli Studi di Bologna - Via Zamboni, 38 - 40126 BOLOGNA. E-mail: buzzetti@philo.unibo.it

mente rappresentato da forme di *markup* non vincolato. La comprensione dello *status* logico del *markup* è comunque essenziale alla sua utilizzazione critica. Il *markup* rende esplicite forme di autoreferenzialità del testo in cui si esprime l'indeterminazione del rapporto tra espressione e contenuto e attraverso il *markup* può essere rappresentata la compensazione dinamica tra la struttura dell'espressione (variante testuale) e la struttura del contenuto (interpretazione del testo).

1. La rappresentazione digitale del testo richiede la definizione di un modello? Ogni forma di rappresentazione in sé e, di conseguenza, ogni forma di rappresentazione del testo comporta l'assunzione implicita o esplicita di un modello, almeno se si accetta il postulato che la «mappa non è il territorio»¹. Sicché la rappresentazione convenzionale di un testo, manoscritto o a stampa, è essa stessa un modello del testo. Altrettanto può dirsi, quindi, della sua rappresentazione digitale o, per essere più precisi, di ogni sua rappresentazione digitale, qualunque ne sia la forma specifica. Il problema del modello viene a porsi, allora, come problema di adeguatezza rispetto alla rappresentazione e al modello convenzionale.

Rispetto alla rappresentazione convenzionale, un'adeguata rappresentazione digitale non dovrebbe comportare alcun impoverimento del contenuto informativo del testo. Se la rappresentazione digitale del testo non è originaria, ossia se si considera un testo riprodotto, anziché prodotto direttamente in forma digitale, un primo fondamentale criterio di adeguatezza è costituito dall'esaustività della rappresentazione. Per ottenere l'esaustività della rappresentazione si ricorre ordinariamente al *markup*. Infatti, il modo di rappresentare informazione di tipo testuale in *machine-readable form* è costituito primariamente dalla codifica binaria dei caratteri: il 'testo' viene così concepito, da un punto di vista informatico, come un tipo di dato e il trattamento del testo, ossia «l'archiviazione e l'elaborazione di materiale testuale», viene a consistere nel trattamento di «informazione codificata come caratteri o come sequenze di caratteri» (Day, 1984: 1). È evidente, però, che

¹ Tale «la proposizione originale per la quale Korzybski va più famoso» e che «scaturì da un vastissimo fiume di riflessione filosofica, che risale alla Grecia e serpeggiò attraverso la storia del pensiero europeo per tutti gli ultimi duemila anni» (Bateson, 1976: 465).

la nozione informatica del testo come tipo di dato non coincide con la nozione del testo inteso come prodotto dell'attività letteraria. La pura e semplice sequenza dei caratteri non è sufficiente a rappresentare tutta l'informazione contenuta nel «materiale letterario qual è originariamente scritto da un autore» (Day, 1984: 1). Di qui la necessità di fornire informazione aggiuntiva mediante l'inserimento di marcatori definiti da un particolare linguaggio di *markup*.

Un secondo criterio di adeguatezza, altrettanto importante, riguarda le possibilità di trattamento automatico della rappresentazione digitale e la sua funzionalità rispetto ad operazioni critiche di costituzione o di interpretazione del testo. Per particolari finalità analitiche, la rappresentazione digitale del testo può presentare vantaggi evidenti ed essere preferibile alla stessa rappresentazione convenzionale. Per esempio, la possibilità di associare immagini digitali alle trascrizioni del testo rende superflua la funzione mimetica delle trascrizioni diplomatiche e ne trasforma le finalità. Associata all'immagine, una trascrizione diplomatica non serve più a «riprodurre la fonte», ma ad estrarne informazioni e a rappresentarle in forma elaborabile automaticamente. In tale prospettiva, i segni diacritici o le forme di *markup* non vengono più concepiti come un ausilio per ricostruire visivamente un documento assente, ma come un mezzo per «modellizzare» l'informazione fisica e testuale contenuta nell'originale ai fini di ulteriori elaborazioni (Buzzetti, 1995: 148). L'immagine stessa, in quanto rappresentazione digitale di informazione visiva, non restituisce solo un «faccsimile» o una «rappresentazione fisica» della fonte, ma offre un insieme di «dati strutturati», ovvero una «rappresentazione logica» del contenuto del documento (Buzzetti, 1995: 148; Kirschenbaum *et al.*, 1999: 56).

La struttura della rappresentazione digitale diviene così rilevante per le condizioni di adeguatezza, poiché impone vincoli precisi ai processi di elaborazione automatica del contenuto informativo del documento. La forma della rappresentazione deve essere funzionale alle operazioni analitiche necessarie allo studio del testo. Il problema si pone già, con tutta evidenza, per la forma di rappresentazione convenzionale del testo (particolari artifici tipografici sono stati proposti, ad esempio, per l'edizione critica di

testi a tradizione testuale fluida o molto complessa) (cfr. Steer, 1993: 118; Buzzetti e Rehbein, 1998: 21ss.) e lo stesso problema si ripropone, con altrettanta se non maggiore evidenza, per la rappresentazione digitale del testo. In ogni caso, non si può assolutamente considerare adeguata una forma di rappresentazione che ponga limiti di principio ad operazioni analitiche praticate dagli studiosi. Un'adeguata rappresentazione digitale del testo deve quindi essere compatibile con l'applicazione dei procedimenti formali di elaborazione dell'informazione che danno forma algoritmica ai diversi metodi di indagine della pratica critica ed interpretativa.

Ora, la pratica del *markup* affermatasi tra gli studiosi che applicano procedure computazionali allo studio del testo ha messo in luce alcune difficoltà dipendenti da fondamentali opzioni teoriche. Tali difficoltà investono rispettivamente entrambe le condizioni di adeguatezza. Da un lato è mancata una piena consapevolezza dello status teorico del *markup* in quanto tale e delle sue diverse forme, con conseguenze negative per l'esaustività della rappresentazione e la pertinenza del *markup*. Dall'altro non sempre si sono ottenute forme di rappresentazione congruenti con le procedure computazionali necessarie all'applicazione di metodi automatici allo studio critico del testo.

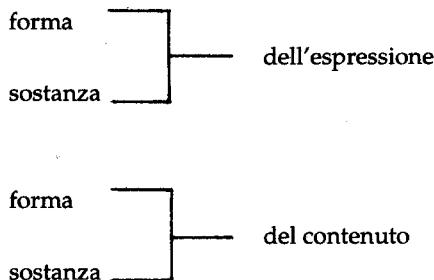
2. Fondamentalmente, il problema dell'adeguatezza della rappresentazione digitale del testo si può formulare nei termini seguenti: (a) quali sono le caratteristiche formali del tipo di dato che realizza la rappresentazione del testo? (b) quali sono le caratteristiche formali del modello strutturale del testo necessarie per implementare le operazioni richieste dalla sua analisi critica? (c) quale formalismo compatibile con la sintassi del dato e con la semantica del modello è possibile implementare? La forma della rappresentazione definita dal tipo di dato dev'essere operativamente congruente con le proprietà strutturali dell'oggetto rappresentato definite dal tipo di modello.

3. Tale impostazione del problema presuppone una precisa distinzione tra struttura della rappresentazione e struttura dell'oggetto rappresentato, una distinzione che non pare sia colta con sufficiente chiarezza

nelle *Guidelines* della Text Encoding Initiative (TEI) (Sperberg-McQueen e Burnard, 1994), la più organica proposta finora avanzata per la «codifica dei testi» e per l'«interscambio dei dati nel campo della ricerca umanistica» (Burnard, 1991: 83). Nella definizione stessa di «che cos'è il *markup*», proposta dai curatori delle *Guidelines* si avverte una «fondamentale ambiguità» (Buzzetti, 1999: 219) che non permette di discriminare tra le proprietà formali della «rappresentazione» dell'informazione e le proprietà formali dell'«informazione» rappresentata (Buzzetti e Rehbein, 1998: 30-31), ovvero tra i dati intesi come «informazione codificata in un certo modo» (Day, 1984: 1) e i dati intesi come il modello o la «struttura astratta» dell'informazione (Buzzetti e Rehbein, 1998: 30). Infatti, se il *markup* viene definito, com'è definito dagli stessi curatori delle *Guidelines* della TEI, come «tutta l'informazione contenuta in un *file* e *diversa* dal testo stesso», com'è possibile che «qualunque *aspetto* del testo che sia di qualche importanza per il ricercatore» possa mai «essere indicato dal *markup*»? (Burnard e Sperberg-McQueen, 1991: 2; corsivi miei). Delle due l'una: o, come anche altri sostengono, il *markup* viene inteso come informazione che «*non fa* parte del testo» (Coombs et al., 1987: 934; corsivo mio) ed è *diversa* dal testo, oppure il *markup* rappresenta certi aspetti dell'informazione che «*fa* parte del testo e *coincide* col testo» (Buzzetti e Rehbein, 1998: 31). Ma certamente non le due cose insieme, a meno che inavvertitamente, come qui pare, *testo* sia inteso in due sensi diversi: da un lato, come una sequenza di caratteri codificati, ossia come espressione o rappresentazione digitale dell'informazione; dall'altro, come «contenuto dell'espressione» o rappresentazione digitale dell'informazione (Coombs et al., 1987: 934). Dovrebbe infatti esser chiaro che «la rappresentazione di un'informazione, o di un contenuto, non è il contenuto o l'informazione rappresentata», o espressa, «da tale rappresentazione» (Buzzetti, 1999: 219).

D'altra parte, le due accezioni, benché distinte, possono essere considerate come complementari ed in effetti debbono essere opportunamente ricomprese entrambe in una nozione adeguata del testo. La teoria del testo «letterario» (cfr. Segre, 1985) presenta talvolta la distinzione tra rappresentazione dell'informazione e informazione rappresentata come distinzione tra «espressione» e «contenuto» del testo (Segre, 1985: 49 ss.). A questo proposito può essere conveniente richiamare la descrizione data da Hjelmslev

(cfr. Hjelmslev, 1968: 52-65), che propone un «modello a quattro piani» di «stampo saussuriano (espressione = significante; contenuto = significato)» (Segre, 1985: 49-50):



Come si vede, questo modello permette di distinguere chiaramente tra la forma della rappresentazione e la forma del contenuto rappresentato.

Un ulteriore tratto saussuriano del modello proposto da Hjelmslev si manifesta «nell'insistenza sull'inscindibilità di espressione e contenuto» (Segre, 1985: 50): da un lato «il testo non può essere identificato con nessuna delle sue diverse forme di rappresentazione», dall'altro la «rappresentazione del testo è necessariamente concreta e richiede un supporto fisico» (Buzzetti e Rehbein, 1998: 29). Sicché non può darsi testo senza rappresentazione; anzi, il testo è rappresentazione, ma rappresentazione che dipende da un contenuto astratto e si determina individualmente nella singolarità di un'espressione concreta. Tutto questo permette di caratterizzare il testo come *invariante*, generalizzando un'osservazione di Cesare Segre riferita alla «costituzione originaria» del testo come «serie di lettere e segni di interpunkzione», che sono portatori di «significati grafici», i quali a loro volta sono «portatori di significati semantici» (Segre, 1985: 29). Ora, se ci riferiamo a questi 'significanti' originari del testo², ossia:

² L'idea del testo come costituito da una 'serie di elementi' che possono essere considerati indipendentemente o come interdipendenti l'uno dall'altro è sviluppata da Tito Orlando nel suo contributo in questo volume: il testo viene così descritto come un 'sistema' composto da una serie di '(sotto)sistemi' e il modello del testo come costituito da un «insieme inscindibile di più livelli» di rappresentazione formale completa e coerente.

se consideriamo i segni grafici (lettere, interpunkzione, 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. (Segre, 1985: 29; corsivo mio).

Se assumiamo come ‘valori’, rispetto a diverse espressioni del testo, i significati semantici di cui esse sono portatrici, possiamo concludere che l’insieme di tali significati costituisce «il contenuto *invariante* di tutte le rappresentazioni materiali» del testo (Buzzetti e Rehbein, 1998: 29).

Dalle stesse considerazioni si può desumere una nozione del ‘contenuto’, o ‘significato’, del testo come classe di equivalenza, sotto particolari riguardi, delle sue diverse espressioni³. Questa nozione si può applicare, nello stesso senso, a singole parole, frasi e passaggi, ossia all’intero testo o a sue singole parti indifferentemente. Ciò permette di rapportare la nozione di variante testuale alla nozione di sinonimia e di stabilire un nesso tra lezioni varianti del testo, o forme varianti dell’espressione, da una parte, e varianti interpretative, o forme varianti del contenuto, dall’altra. La rilevanza di queste considerazioni per l’analisi della rappresentazione digitale del testo verrà discussa più diffusamente in seguito (cfr. *infra*, sez. 12). Qui importa soltanto mettere in luce, da un punto di vista astratto, la correlazione tra la forma dell’espressione e la forma del contenuto del testo, una correlazione che non comporta tuttavia, in generale, una necessaria condizione di isomorfismo. Il carattere concomitante e complementare della forma dell’espressione e della forma del contenuto costituisce senz’altro un requisito necessario del testo, ma lo stesso non si può dire del loro isomorfismo. Al contrario, occorre ribadire che questa circostanza si realizza solo in condizioni molto particolari.

4. La distinzione tra ‘espressione’ e ‘contenuto’ può essere opportunamente applicata anche alla rappresentazione digitale del testo. In forma digitale, il testo può essere considerato come un ‘dato’ e a proposito di dati

³ Per un punto di vista analogo, cfr. Naess (1953). Debbo la segnalazione a Claus Huitfeldt nel corso di conversazioni avute su questi temi presso i ‘Wittgenstein Archives’ di Bergen.

è possibile distinguere tra ‘rappresentazione di dati’ e ‘modello di dati’⁴. Il dato può essere definito «in informatica» come «rappresentazione (di informazione) in un certo codice» (La Torre, 1994: 333) dove per ‘rappresentazione’ si intende un insieme di segni, ovvero un insieme di «oggetti», costituiti dagli stati fisici «su cui operano gli esecutori di algoritmi», cioè gli elaboratori; e dove per ‘informazioni’ si intendono i «significati» che sono «attribuiti ai dati», ossia all’insieme dei segni, degli oggetti, o degli stati fisici che ne costituiscono l’espressione materiale (La Torre, 1994: 217). Ora, la forma, o l’organizzazione degli oggetti che costituiscono la rappresentazione concreta dell’informazione, in breve, la *forma dell’espressione* dell’informazione, è ciò che viene correntemente chiamato una ‘struttura di dati’, mentre la *forma del contenuto* dell’informazione, la struttura o l’organizzazione del contenuto informativo, è ciò che viene correntemente chiamato un ‘modello di dati’⁵. Sicché anche la nozione di ‘dato’, come la nozione di ‘testo’, è una nozione ambigua e può riferirsi sia all’espressione, sia al contenuto dell’informazione – una constatazione che non dovrebbe stupire, se si considera che la rappresentazione digitale di un testo è per definizione un insieme di dati.

Si deve necessariamente tener conto di questa doppia dimensione dei dati, l’espressione e il contenuto, quando si riflette sulla natura e sullo status logico del *markup*. Un altro modo di esprimere la distinzione tra espressione e contenuto, con riferimento ai dati, è quello di distinguere tra ‘formato’ e ‘formalismo’; e anche a questo proposito è stato rilevato che, «sebbene i due termini (formato e formalismo) abbiano significato diverso, spesso vengono confusi». Ora, un *formato* può essere definito come la «sintassi» secondo cui una certa «rappresentazione» dell’informazione è stata «codificata», mentre un *formalismo* può essere descritto come ciò che fornisce un’«interpretazione» a tale «rappresentazione» dell’informazione, ossia come ciò che assegna un certo «significato» alla «sequenza dei valori digitali» di cui è costituita (Joloboff, 1989: 75-6). Il ‘formato’ della rappresentazione è ciò che determina le ‘strutture di dati’, ossia le forme strutturali delle espressioni di cui è costituita, il ‘formalismo’ invece è ciò che deter-

⁴ Cfr. Raymond *et al.* (1992), <URL: <http://www.csd.uwo.ca/staff/drraymon/papers/markup-ps>>.

⁵ È in questo senso che vengono usate, rispettivamente, le due nozioni da Raymond *et al.* (1992: *passim*).

mina il ‘modello di dati’, ossia la «forma astratta» (Raymond *et al.*, 1992: 13) del contenuto rappresentato da tali espressioni. ‘Formato’ e ‘struttura di dati’ sono nozioni sintattiche e si riferiscono all’espressione della rappresentazione digitale; ‘formalismo’ e ‘modello di dati’ sono nozioni semantiche e si riferiscono al suo contenuto. Un modello di dati «deve rappresentare conoscenze» di fatto (Goldschlager e Lister, 1988: 248) sugli oggetti che i dati rappresentano e consiste nella specificazione formale delle proprietà astratte degli oggetti rappresentati, insieme con le operazioni definibili su tali elementi e le restrizioni che si applicano alle diverse classi di oggetti. Un modello di dati può quindi definire un’«algebra» per i suoi oggetti⁶, valendosi di un formalismo matematico «in modo» affatto «indipendente dalla» particolare forma di «rappresentazione dei dati» (Herstein, 1975: 4-5).

5. Tenendo conto di queste distinzioni, è ora possibile cercare di descrivere in modo adeguato la natura propria del *markup*. È stato osservato che «il *markup* è semplicemente la denotazione di specifiche posizioni» nella sequenza dei caratteri che costituiscono l’espressione del testo mediante l’inserimento di certi «contrassegni (*tokens*)», costituiti essi stessi da caratteri (Raymond *et al.*, 1992: 4). Il *markup* può quindi essere definito, in generale, come «l’uso di codici, detti marcatori (*tags*), inseriti in un documento, per descriverne la struttura» (Raymond *et al.*, 1992: 1). Si tratta quindi di «una delle diverse tecniche possibili per rappresentare la struttura» di un «documento», ossia la struttura di ciò che costituisce propriamente l’espressione del testo (Raymond *et al.*, 1992: Abstract). Si può così sostenere che «il *markup* non appartiene al mondo dei formalismi, ma al mondo delle rappresentazioni» (Raymond *et al.*, 1992: 4) oppure che «il *markup* non è un modello di dati», ma «un tipo di rappresentazione dei dati» (Raymond *et al.*, 1992: 16). Infatti, mentre un modello di dati descrive «direttamente la semantica, senza prendere in considerazione la rappresentazione dei dati» (Raymond *et al.*, 1995: 5), è evidente che «i problemi del *markup* sono problemi che riguardano la rappresentazione della struttura e non la sua forma

⁶ Raymond *et al.* (1995: 15), <URL: <http://www.cs.d.uwo.ca/staff/draymond/papers/sgml.ps>>. Per ‘algebra’ si intende qui un tipo di struttura matematica che può essere genericamente descritto come «un aggregato di oggetti insieme con certe operazioni per combinarli» (Herstein, 1975: 2).

astratta» considerata indipendentemente dall'espressione (Raymond *et al.*, 1992: 13; corsivo mio).

Si possono tuttavia distinguere, fondamentalmente, due diversi modi di assegnare ad un documento una struttura e, di conseguenza, due diversi tipi di *markup*. Infatti la struttura può essere «inserita» o «incastrata» nel testo, inteso come espressione o stringa di caratteri codificati, «in modo vincolato o in modo sciolto (*strongly or weakly embedded*)». In entrambi i casi i marcatori che denotano la struttura sono contenuti nella rappresentazione dei dati. Però, nel primo caso «la posizione» in cui essi sono inseriti nella sequenza dei caratteri «è portatrice di informazione»; invece nel secondo caso, benché i marcatori siano «informativi», la loro «collocazione all'interno del testo», o della sequenza dei caratteri, «non è portatrice di informazione». Si può così parlare, rispettivamente, di *markup* '(inserito in modo) vincolato (*strongly embedded*)' e di *markup* '(inserito in modo) sciolto (*weakly embedded*)' o 'non vincolato' (Raymond *et al.*, 1992: 3-4).

È evidente allora, come è stato fatto osservare, che le «proprietà» delle strutture che possono essere descritte da una forma di *markup* vincolato «sono in gran parte derivate dalle proprietà del documento», o della stringa di caratteri, «in cui esso viene inserito» (Raymond *et al.*, 1992: 4). Infatti, «poiché condivide coi dati la forma di rappresentazione», diventa «difficile» per il *markup* vincolato «esprimere una struttura che non sia un sottoinsieme di posizioni di caratteri nel testo» (Raymond *et al.*, 1992: 9). Il fatto di essere indissolubilmente connesso con la posizione in cui viene inserito, fa sì che il *markup* vincolato «erediti molte delle proprietà del testo» inteso come sequenza di caratteri (Raymond *et al.*, 1992: 2) e in primo luogo il suo «ordine» lineare (Raymond *et al.*, 1992: 9, 10). Ma poiché «una struttura di alto livello non è sempre composizione di caratteristiche di livello inferiore», la struttura del testo «non è sempre riducibile ad una descrizione funzionale di sottocomponenti del sistema», quali la posizione relativa dei caratteri nel loro ordinamento lineare (Raymond *et al.*, 1992: 9). Il *markup* vincolato non pare quindi del tutto adeguato ad esprimere la struttura del testo e i suoi «problemi» possono essere considerati il «risultato» delle «proprietà che eredita» dalla sequenza dei caratteri (Raymond *et al.*, 1992: 2).

Le stesse cose possono essere dette in altro modo facendo riferimento alla distinzione fra l'*espressione* e il *contenuto* dei dati e considerando, rispettivamente, la loro *forma* specifica, ossia l'ordine lineare della successione dei caratteri codificati da un lato e la struttura di ciò che le diverse stringhe di caratteri significano dall'altro. Si può così affermare che la dipendenza del *markup* vincolato dalla *forma dell'espressione* dei dati limita gravemente le sue capacità di rappresentazione della struttura del testo. La *forma del contenuto* dei dati non può essere descritta, in generale, da una funzione della posizione di determinate stringhe di caratteri all'interno della sequenza lineare che le contiene. Infatti, la forma del contenuto dei dati non è necessariamente isomorfa alla forma della loro espressione.

In sostanza, quindi, un linguaggio di *markup* vincolato è un linguaggio per rappresentare *strutture di dati* e non un linguaggio per rappresentare *modelli di dati*. Un linguaggio di *markup* vincolato si riferisce alla rappresentazione dell'informazione e non al contenuto dei dati; serve per descrivere la forma dell'espressione dei dati e non «per descrivere restrizioni» di fatto riguardanti gli elementi strutturali dell'informazione rappresentata o per specificare «gli effetti delle operazioni» che si possono eseguire combinando tali oggetti formali (Raymond *et al.*, 1995: 6). Un sistema di *markup* vincolato può quindi essere descritto come una forma di rappresentazione del *formato* dei dati e non del *formalismo* che eventualmente definisce gli operatori applicabili a quei dati. In ultima analisi, esso serve fondamentalmente per esprimere proprietà strutturali riguardanti le caratteristiche 'notazionali' della rappresentazione dell'informazione.

6. Il *markup* vincolato è dunque essenzialmente una rappresentazione della *forma dell'espressione* dei dati e serve ad esplicitare la struttura della rappresentazione del testo. Esso rappresenta la rappresentazione del testo e segnatamente la sua struttura. Occorrerà riflettere più a fondo sullo status del *markup* in quanto *rappresentazione della rappresentazione* del testo (cfr. *infra*, sez. 12); piuttosto, se è vero che «le proprietà strutturali di un documento codificato espresse da una forma di *markup* vincolato» sono «essenzialmente proprietà 'notazionali'» (Buzzetti, 1999: 220), ora mette conto ribadire che le proprietà notazionali della rappresentazione del testo, essendo proprietà della forma dell'espressione dei dati, non debbono in alcun modo essere confuse con le proprietà del loro contenuto. A questo pro-

posito si può osservare, per analogia, che le differenze di notazione tra la numerazione decimale, la numerazione binaria, o qualunque altro sistema di numerazione non si riflettono in alcun modo sulle proprietà aritmetiche dei numeri naturali. L'operazione descritta qui di seguito è sempre la stessa indipendentemente dal sistema di notazione adottato, decimale o binario, arabo o romano che sia (fig. 1).

decimale	4 + 2	= 6	arabo
binario	100 + 10	= 110	arabo
decimale	IV più II è uguale a VI		romano

Fig. 1.

Allo stesso modo, la struttura lineare dell'espressione di un'equazione di secondo grado usata per descrivere una sezione conica è del tutto indipendente dalle proprietà geometriche della figura rappresentata (fig. 2).

Questi esempi mostrano con tutta evidenza che è assolutamente improprio far dipendere le proprietà strutturali del contenuto di una rappresentazione dalle forme strutturali che la sintassi del linguaggio di rappresentazione assegna alla sua espressione. Sicché quello che un linguaggio di *markup* permette di esplicitare è solo la *forma dell'espressione* del testo, ovvero la struttura logica di un documento. Inoltre, la struttura assegnata al documento dipende dalle caratteristiche sintattiche del linguaggio di *markup* usato per rappresentarla. Infatti, «non esiste una sola struttura logica per lo stesso documento» (Raymond *et al.*, 1992: 16) e la forma dell'espressione del testo di volta in volta rappresentata dipende dalle capacità espressive del sistema di *markup* vincolato che permette di esplicitarla. Di conseguenza, non è legittimo identificare la struttura del *contenuto* del testo con la struttura dell'*espressione* del testo che un particolare linguaggio di *markup* è in grado di rappresentare.

7. Espressione e contenuto del testo vengono tuttavia confusi, come si è visto, nelle *Guidelines* della TEI e nella pratica di codifica che ne consegue. Responsabile della confusione sembra essere stata l'ambigua definizione del testo «come 'gerarchia ordinata di oggetti dotati di contenuto' ovvero 'OHC' (Ordered Hierarchy of Content Objects)» e l'affrettata assunzione di

questa definizione come «fondamentale modello del testo» (De Rose *et al.*, 1990: 6). Un ‘oggetto dotato di contenuto’ è una porzione di «documento» (De Rose *et al.*, 1990: 5) che contiene o è contenuta in altri oggetti dotati di contenuto, o porzioni del documento, e che forma con essi una ‘gerarchia’ di relazioni di contenimento, i cui elementi minimi sono ‘ordinati’ in successione nella sequenza di caratteri di cui il documento si compone. La nozione di ‘oggetto dotato di contenuto’ permette in sostanza di concepire un documento come una struttura di dati e segnatamente come un grafo ad albero i cui nodi estremi sono ordinati linearmente.

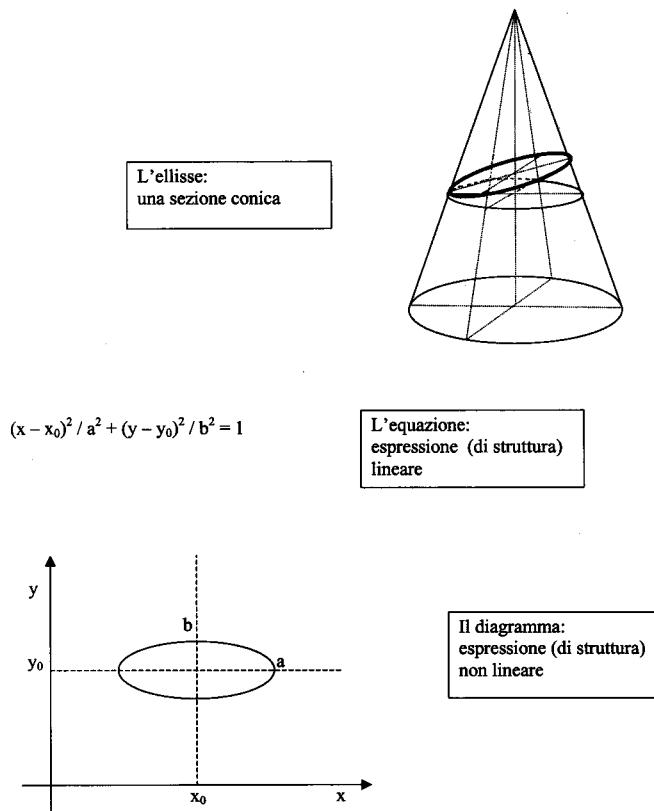


Fig. 2.

Questa definizione è stata implicitamente assunta nella pratica ordinaria del lavoro di codifica con la «promozione di SGML», lo standard ISO

per sistemi di *markup* descrittivo⁷, «a standard per la codifica di dati testuali» (De Rose *et al.*, 1990: 18). Infatti la TEI si è «affidata all’uso di SGML come linguaggio fondamentale per la descrizione del testo» (De Rose *et al.*, 1990: 13) e «SGML definisce un documento come una struttura OHCO»: ciò che esso rappresenta è «una struttura gerarchica del documento» descritta «con nomi mnemonici per gli oggetti dotati di contenuto che costituiscono i dati» di cui si compone (De Rose *et al.*, 1990: 12). In sostanza,

una premessa fondamentale di SGML assume che i testi siano composti da oggetti discreti dotati di contenuto (*discrete content objects*) e che l’assegnazione di nomi significativi a questi oggetti testuali ben delimitati, ai loro attributi e alle loro relazioni gerarchiche sia uno dei mezzi più efficaci per trasformare il testo [il flusso di caratteri] in unità di informazione che possono essere individuate sensatamente da programmi in grado di discriminare (Cover *et al.*, 1981: 198).

È evidente che questo presupposto fondamentale assimila espressione e contenuto del testo e confonde la struttura determinata dagli oggetti dotati di contenuto, una struttura formata da porzioni del documento costituite da stringhe di caratteri, con la «struttura del contenuto» del documento (Cover *et al.*, 1981: 23), una struttura formata da elementi astratti di oggetti di conoscenza. Infatti essa si fonda sulla «tesi» che, «in una certa accezione dei termini ‘libro’, o ‘testo’, o ‘documento’ (forse in quanto *oggetti intellettuali*), tutte queste cose siano ‘gerarchie ordinate di oggetti dotati di contenuto’» (Renear *et al.*, 1995: 265), dove è palese l’assimilazione del testo al libro o al documento che ne realizza materialmente la rappresentazione.

Ora, non è SGML responsabile di questa confusione, ma l’uso che ne è stato fatto per rappresentare il contenuto del testo anziché la sua espressione. SGML viene descritto come un «metalinguaggio» (Smith, 1992: 15), che «non definisce un linguaggio di *markup*», ma la sintassi formale per l’uso di marcatori vincolati. SGML è quindi «un linguaggio per definire *linguaggi di markup*» di tipo vincolato (Burnard e Sperberg-McQueen, 1991: 2)

⁷ Cfr. ISO 8879:1986, Information processing – Text and office systems – Standard Generalized Markup Language (SGML). Il *markup* descrittivo può essere definito come l’uso di “nomi mnemonici” per gli oggetti dotati di contenuto (De Rose *et al.*, 1990: 5) o più in generale per i diversi elementi della rappresentazione del testo.

che consentono l'esplicitazione della struttura, o forma, dell'espressione dei dati in termini di posizioni relative all'ordinamento lineare dei caratteri codificati; in breve, SGML è la specificazione formale della sintassi dell'espressione dei dati. Dunque, SGML è un «linguaggio di rappresentazione» concepito per descrivere «la struttura logica di un documento» (Smith, 1992: 15), cioè la struttura dell'espressione del testo. La TEI invece «assume implicitamente questo modello strutturale come modello fondamentale di rappresentazione del testo e vincola di conseguenza la forma di rappresentazione del testo ad una struttura di dati» che descrive di fatto la struttura della sua espressione (Buzzetti, 1996a: 91). Una 'struttura OHCO' non è un modello del testo, ma un possibile modello della sua espressione.

La «tesi OHCO», tuttavia, non è più accettata senza riserve. Anzi, l'attività stessa di codifica dei testi ha fatto emergere «alcuni problemi pratici che sembrano mettere questa tesi in discussione» (Renear *et al.*, 1995: 263). In sostanza, «il modo in cui i testi venivano risolti in oggetti dai teorici dell'elaborazione del testo (*text processing*) e dagli sviluppatori di standard» di codifica, questo modo fondato sulla nozione di 'testo' come flusso di caratteri, si è rivelato «radicalmente diverso dal modo in cui i testi sono risolti in oggetti dalla comunità degli studiosi di fenomeni letterari e linguistici». Sicché «la tendenza di SGML ad assumere che i documenti possano essere rappresentati da un'unica struttura gerarchica ha creato seri problemi pratici ai progetti di codifica dei testi». Ma non si tratta solo di concludere che la 'tesi OHCO' «è falsa» perché si possono assegnare «più strutture gerarchiche» ad uno stesso documento, tutte aventi titolo per «reclamare plausibilmente di essere la struttura 'logica'» che gli compete (Renear *et al.*, 1995: 269 ss.), bensì di ammettere che «le strutture logiche rilevanti di un certo testo non sono tutte senza eccezione di tipo gerarchico» (Renear *et al.*, 1995: 279, nota 13). Infatti, non solo si deve convenire che «lo stesso documento si conforma a diverse strutture sovrapposte» (Barnard *et al.*, 1988: 266), ciascuna «strettamente gerarchica», cioè composta di «oggetti sempre 'annidati' o 'contenuti l'uno nell'altro' (*nested*) e mai 'sovraposti' (*overlapped*)» (Renear *et al.*, 1995: 269), ma si deve riconoscere che le strutture testuali non sono in generale di questo tipo e che ogni «struttura non gerarchica costituisce un problema per un sistema di codifica dei testi letterari fondato su SGML» (Barnard *et al.*, 1988: 266-67). SGML è un linguaggio di rappresentazione di strutture di dati e il problema nasce chiaramente dalla

confusione fra la struttura dell'espressione del testo che può essere rappresentata con un *markup* di tipo vincolato e la struttura del contenuto del testo che non si conforma, in generale, ad un modello di tipo gerarchico e lineare. SGML, inoltre, non è un linguaggio di rappresentazione di modelli di dati; infatti, «non definisce nessun operatore» (Raymond *et al.*, 1995: 6, 15) e non fornisce «nessuna semantica» (Raymond *et al.*, 1992: 17). L'uso di sistemi di *markup* vincolato pone dunque, in generale, serie limitazioni all'adeguatezza delle rappresentazioni del testo: l'impiego di «una rappresentazione lineare universale di strutture gerarchiche» è «un'imposizione che impedisce drasticamente la rappresentazione di strutture non gerarchiche» (Nelson, 1977: 129-30).

Ora, è vero che «qualsiasi struttura di dati», lineare e non lineare, «può essere rappresentata» linearmente da un «flusso» di caratteri «conforme» alla sintassi formale di SGML (Raymond *et al.*, 1995: 6), che è «una tecnica per la rappresentazione sequenziale e gerarchica di ogni tipo di dato con marcatori vincolati (*embedded*) che rappresentano posizioni iniziali e finali» (Nelson, 1977: 129); questo però non significa che tale struttura «possa essere inserita in modo vincolato (*strongly embedded*) nella sequenza lineare dei caratteri» (Buzzetti e Rehbein, 1998: 34) e possa essere rappresentata esclusivamente in funzione della posizione dei marcatori al suo interno. La struttura sintattica che SGML assegna alla rappresentazione è in grado cioè di esprimere non tale struttura, ma la forma, o struttura, della sua rappresentazione. Infatti, una struttura non-lineare può essere rappresentata linearmente, ma non può essere *definita* linearmente, ossia calcolata esclusivamente in funzione di un ordine puramente lineare qual è quello della posizione dei caratteri nella sequenza lineare che costituisce l'espressione del testo.

In conclusione, si può quindi affermare che le forme di *markup* vincolato risultano inadeguate sia dal punto di vista dell'esaustività, sia dal punto di vista della funzionalità della rappresentazione del testo e del suo modello. Le forme di *markup* vincolato non sono in grado di rappresentare strutture non lineari in funzione di posizioni assegnate all'interno della sequenza di caratteri che costituisce l'espressione del testo e quindi non soddisfano il criterio dell'esaustività della rappresentazione. In ultima analisi, il *markup* vincolato fa collassare la forma, possibilmente non lineare, del con-

tenuto dei dati sulla forma, esclusivamente lineare, dell'espressione dei dati. La linearizzazione dell'*espressione* non può tuttavia comportare la linearizzazione del contenuto e della sua struttura. Quindi il *markup* vincolato non è in grado di rappresentare strutture non lineari del *contenuto* dei dati. Ma non permette nemmeno di ottenere rappresentazioni del testo funzionalmente adeguate. Infatti, la rappresentazione lineare di una struttura non lineare di dati, il formato dell'espressione dei dati ottenuto con forme di *markup* vincolato, non è solitamente associato alla definizione di operatori applicabili agli elementi della rappresentazione. L'assenza di definizioni operative associate alla rappresentazione del testo non rende di per sé possibile, senza ulteriori specificazioni, l'applicazione di procedure algoritmiche definite da un coerente modello di dati agli elementi e alle strutture di dati che essa distingue. È certo possibile associare un formalismo ad «un sistema» di gestione di documenti «che riconosca testi [stringhe di caratteri] contenenti una struttura (solitamente indicata dalla presenza del *markup*)»; tuttavia non può trattarsi altro che di un «formalismo (solitamente una grammatica libera dal contesto)» (Raymond *et al.*, 1992: 4) idoneo al trattamento di strutture di dati lineari, come ad esempio le 'strutture OHCO'. Di conseguenza, le forme di *markup* vincolato non riescono a soddisfare, in generale, il criterio della funzionalità della rappresentazione.

8. Diversamente dal *markup* vincolato, il *markup* 'non vincolato' o 'inserito in modo sciolto (*weakly embedded*)' non pone limiti a priori al requisito dell'adeguatezza funzionale della rappresentazione del testo. Per *markup* non vincolato si intende «una struttura non inserita (*embedded*) [nel flusso di caratteri] che si conforma ai requisiti sintattici di un certo standard di *markup*». Un «marcatore» non vincolato, o «inserito in modo sciolto (*weakly embedded*)», è quindi «portatore di informazione», ma «la sua posizione» all'interno della sequenza di caratteri «non è informativa». Esso può essere collocato «in qualsiasi punto all'interno del testo [flusso di caratteri], o anche al di fuori del testo»; per questa ragione viene anche chiamato *markup* 'fuori linea' (*out-of-line*) ed è «più propriamente considerato come una forma specifica di struttura esterna». È quindi possibile «distinguere tra struttura interna e struttura esterna» al 'testo', ma dev'essere ben chiaro che ciò a cui qui si fa propriamente riferimento è la sequenza lineare dei caratteri che costituisce l'*espressione* del testo (Raymond *et al.*, 1992: 3-4). Le ragioni per cui una struttura 'interna' all'espressione del testo è isomorfa al

suo ordinamento lineare non può corrispondere, in generale, alla struttura del suo contenuto sono state sufficientemente chiarite. Di conseguenza solo una struttura 'esterna' all'espressione del testo può essere in grado di fornire un modello adeguato del suo contenuto.

Una struttura esterna all'espressione del testo costituisce di solito un modello di dati. Il riferimento ad un modello di dati capovolge, in sostanza, la relazione tra le strutture di dati definite dalla rappresentazione del testo e il particolare formalismo che ne regola la manipolazione. Non è più la struttura assegnata dal *markup* all'espressione del testo ciò che determina le operazioni applicabili ai dati testuali, ma sono le operazioni applicabili agli elementi strutturali del contenuto del testo ciò da cui dipendono le strutture di dati che ne esprimono la rappresentazione. In questo modo un documento, o in generale l'espressione del testo, vengono generati da un modello formale che ne elabora il contenuto informativo; in breve, il documento o l'espressione del testo dipendono dal modello formale che li produce. In ciò si esprime anche un concetto di edizione digitale che si richiama fondamentalmente ad una concezione generativa della rappresentazione del testo.

La soluzione del *markup* esterno o «parallelo» al «flusso» di caratteri (Nelson, 1977: 131) è stata decisamente propugnata, contro la pratica del *markup* vincolato, da Theodor Nelson, il clamato inventore del «termine 'ipertesto'», usato «per esprimere l'idea della scrittura/lettura non lineare in un sistema informatico» (Lévy, 1990: 34). Secondo Nelson, le strutture di dati sequenziali e gerarchiche prescritte dal *markup* vincolato sono assolutamente «inabili» e «frappongono ostacoli» alla «esatta rappresentazione» di «quel tipo di pensiero che viene espresso con le parole e la scrittura» (Nelson, 1977: 132-33). Una rappresentazione adeguata del contenuto del testo può dunque essere ottenuta solo con forme di *markup* parallelo, i cui «marcatori possono» sì «essere simili ai marcatori SGML», ma «non sono» certo «inseriti in modo vincolato (*embedded*)» nel testo, bensì «trattati separatamente» e collocati «in flussi paralleli che fanno riferimento alle posizioni nel flusso dei dati testuali» (Nelson, 1977: 131).

Motivazioni analoghe si ritrovano nella proposta di Manfred Thaller per la rappresentazione e il trattamento del «testo storico» (Thaller, 1993: 61

ss.). Un testo storico è descritto come «la rappresentazione elaborabile formalmente dello stato corrente delle ipotesi di un ricercatore su ciò che i suoi documenti effettivamente contengono» (Thaller, 1993: 63) e presuppone il riferimento ad una struttura esterna che ne organizza il contenuto. Tale struttura deve essere «incorporata (*built in*)» nella rappresentazione «naturale» del testo (Thaller, 1993: 83) intesa come un «gruppo ordinato linearmente di stringhe» di caratteri (Thaller, 1993: 65). Una rappresentazione adeguata del testo deve così poter «definire il rapporto tra il 'testo' considerato come rappresentazione lineare (*running representation*) di un documento tradi-to e il 'testo' convertito in un *database* secondo un certo modello astratto» di dati (Thaller, 1993: 64). Capovolgendo la prospettiva e ponendosi dal punto di vista del modello di dati in grado di rappresentare la struttura del contenuto, una rappresentazione lineare codificata del testo appare come una «realizzazione» (Thaller, 1993: 65) o «rappresentazione esterna» della «rappresentazione interna» al modello (Thaller, 1996: 252). Ora, il modello proposto prevede un «tipo di dato non lineare» (Thaller, 1993: 61) costituito da «espansioni di stringhe (*extended strings*)» (Thaller, 1996: 252 ss.), formate da insiemi ordinati di «contrassegni (*tokens*)» a cui «può essere associato un insieme arbitrariamente complesso di attributi». A sua volta un «contrassegno (*token*)», che ordinariamente è formato da una sola stringa di caratteri, può essere costituito da «un insieme non ordinato di stringhe» (p. es. tutte le forme associate ad un certo lemma) (Thaller, 1999: 139). Tale tipo di dato permette di rappresentare contemporaneamente in forma non lineare diverse strutture gerarchiche sovrapposte, ciascuna delle quali può essere esportata o importata, come stringa di caratteri codificata, da un sistema che implementa tale modello. Più strutture gerarchiche o non gerarchiche possono così essere rappresentate contemporaneamente dal sistema e visualizzate distintamente mediante diverse rappresentazioni lineari di stringhe codificate.

Entrambe queste soluzioni richiamano l'idea dei documenti «virtuali»⁸ o «attivi», che ancora non molto tempo fa erano annoverati tra le «applicazioni future» (Raymond *et al.*, 1992: 9), ma che trovano ormai, con

⁸ Cfr. HTF/VD1: *Workshop on Virtual Documents, Hypertext Functionality and the Web*, organized by M. Miroslavjevic, F. Vitali, and C. Watters, at WWW8: *The Eighth International WorldWideWeb Conference* (Tuesday, May 11, 1999). Toronto, Canada, <URL: <http://www.cs.unibo.it/~fabio/VD99/index.html>>.

lo sviluppo del *WorldWideWeb*, implementazioni sempre più frequenti. Un documento virtuale può essere considerato come un documento prodotto «dinamicamente» con diverse «tecniche di generazione» da un insieme «sottostante» di «dati» o «basi di conoscenza»⁹, ovvero come «un documento per il quale non esiste uno stato permanente e per il quale alcuni o tutti gli esemplari sono generati all'istante»¹⁰. È interessante osservare, a questo proposito, che è stata dedicata particolare attenzione a due forme specificamente distinte di documenti virtuali, i «documenti ipertestuali» e i «documenti concettuali»¹¹, che corrispondono, a grandi linee, ai due tipi di soluzione precedentemente illustrati e che rimandano significativamente a due diversi tipi di edizioni digitali, le edizioni ipertestuali¹² e le edizioni in forma di *database*¹³, di cui esistono già proposte teoriche e realizzazioni concrete.

9. Le considerazioni fin qui svolte sulla natura e sulle diverse forme di *markup* hanno mostrato che un modello di dati adeguato alla rappresentazione del testo è associabile alla sua espressione lineare solo attraverso forme di *markup* esterno e non vincolato. Il *markup*, di qualunque forma esso sia, è essenzialmente notazionale, in quanto associa una struttura all'*espressione* del testo e non è necessariamente connesso con nessun formalismo. Inoltre, le forme di *markup* vincolato generano strutture di dati a cui non sono applicabili formalismi idonei al trattamento di strutture di oggetti di natura non lineare. La struttura assegnabile alla rappresentazione lineare

⁹ Miroslavjevic *et al.* (1999), <URL: <http://www.cs.unibo.it/~fabio/VD99/intro.html>>.

¹⁰ Watters e Shepherd (1999), <URL: <http://www.cs.unibo.it/~fabio/VD99/shepherd/shepherd.html>>.

¹¹ Ranwez e Crampes (1999), <URL: <http://www.cs.unibo.it/~fabio/VD99/ranwez/ranwez.html>>: un “documento ipertestuale” è identificato con un documento virtuale generato dalla ‘navigazione’ dell’utente la cui “forma finale dipende interamente dalle [sue] intenzioni”, mentre un “documento concettuale” è descritto come un documento virtuale costruito dinamicamente da un “motore” che ne sceglie, organizza ed assembla gli elementi informativi (*information bricks*).

¹² Cfr. McGann (1997: 19-46) e *Imagining what you don't know: The theoretical goals of the Rossetti archive*, <URL: <http://www.iath.virginia.edu/~jjm2f/chum.html>>.

¹³ Cfr. Buzzetti e Tabarroni (1991: 185-193), dove viene sviluppata un’idea di ‘edizione come *database*’ a partire dal concetto di ‘*database* come edizione’ introdotto da Manfred Thaller (cfr. Thaller, 1991).

del testo da un sistema di *markup* vincolato dipende dalla forma dell'espressione del testo che non è isomorfa, né coincide, in generale, con la forma del suo contenuto. Il tentativo di assumere il modello di elaborazione che può essere applicato all'espressione della rappresentazione digitale del testo come modello di elaborazione applicabile al suo contenuto genera equivoci e difficoltà insolubili¹⁴. L'elaborazione dell'espressione e l'elaborazione del contenuto richiedono, rispettivamente, due formalismi diversi. La struttura dell'espressione è definita infatti da un ordine lineare, mentre la struttura del contenuto è definita, in generale, da una matrice a più dimensioni.

Per realizzare la congruenza fra la forma dell'espressione e le proprietà strutturali del contenuto rappresentato, occorre individuare un formalismo che sia implementabile secondo la sintassi definita dai tipi e dalle strutture di dati di cui è costituita la rappresentazione lineare del testo e che operi secondo la semantica ad essa assegnata dal modello di dati che ne struttura il contenuto. Questa condizione si realizza con l'impiego di forme di *markup* non vincolato in grado di associare la struttura esterna, generalmente non lineare, del modello del contenuto al flusso, essenzialmente lineare, dell'espressione del testo. Così, per esempio, per superare il problema costituito dalla presenza nel testo di «strutture gerarchiche sovrapposte (*overlapping hierarchies*)» sono state suggerite soluzioni che cercano di «fornire non solo un formato di *file* adatto alla registrazione delle sovrapposizioni», cioè una forma idonea di *markup*, «ma anche una notazione per esprimere restrizioni (*constraints*) valide per i documenti con sovrapposizioni», ossia un formalismo operante sugli oggetti rappresentati, «e strutture di dati plausibili per rappresentare documenti con sovrapposizioni», vale a dire una sintassi per assegnare una struttura adeguata all'espressione del testo (Sperberg-McQueen e Huitfeldt, 1999: 198). Sicché, in conclusione, una rappresentazione digitale adeguata del testo sembra richiedere una forma di *markup* non vincolato e un modello di dati non lineare. Il sistema di *markup* dev'essere in grado di progettare la struttura non lineare del modello sull'espressione lineare del testo e il modello dev'essere in grado di rappresentare la struttura non lineare del suo contenuto. Il modello di dati

14 Riferendoci allo schema illustrato da Tito Orlando nel contributo a questo volume, potremmo dire che si confonderebbero così due 'livelli' diversi del modello del testo, corrispondenti a due diversi 'sotto-sistemi' che richiedono formalismi e implementazioni diverse.

può essere così concepito come struttura profonda, capace di generare distinte strutture superficiali o rappresentazioni lineari del testo, tutte tra loro interconnesse e tutte collegate insieme alla stessa struttura profonda.

Il modello del testo implicitamente presupposto da qualsiasi rappresentazione digitale coincide in effetti col modello di dati impiegato per generarla. Ogni rappresentazione del testo ne presuppone un modello e l'espressione del testo può essere considerata come la struttura algoritmica che ne permette l'elaborazione. Il testo, ogni testo, considerando la struttura materiale che concretamente lo realizza, può essere concepito come un sistema di elaborazione del suo contenuto concettuale. La tecnologia alfabetica non è che un esempio di tale realizzazione. La tecnologia digitale offre migliori possibilità di formalizzazione delle procedure.

10. Se dunque i requisiti che assicurano l'esaustività della rappresentazione digitale del testo risiedono principalmente nelle sue caratteristiche strutturali e formali, i requisiti fondamentali per garantirne l'adeguatezza funzionale debbono essere assicurati dalla definizione di un modello di dati in grado di formalizzare i procedimenti analitici e critici applicabili allo studio del testo. Il modello di dati su cui fondare la rappresentazione digitale del testo dev'essere in grado di trasporre su base algoritmica i procedimenti di critica testuale e di analisi interpretativa del testo. Il modello dev'essere in grado di soddisfare tanto le esigenze del filologo e dell'editore, quanto quelle dello storico e del critico letterario.

A ben vedere, entrambi i procedimenti si fondano su una relazione uno/molti. L'editore infatti deve

riportare ad una rappresentazione unica, coerente e non sequenziale le diverse rappresentazioni sequenziali del testo ricavate dai suoi singoli testimoni. In tal modo l'intera struttura delle relazioni intertestuali tra le diverse redazioni del testo, solitamente rappresentata mediante l'apparato, può essere ricostruita, elaborata ed analizzata con un dispositivo in grado organizzare in strutture di tipo diverso, sequenziali e non sequenziali,

l'intera tradizione testuale, ossia i dati di fatto a disposizione dell'editore per la ricostruzione del testo. Ma «anche a proposito dell'interpretazione, si

tratta di riportare ad una forma di rappresentazione unitaria e coerente una pluralità di diverse rappresentazioni strutturali»; si tratta cioè di

far coesistere una molteplicità di forme strutturali diversamente articolate e interconnesse fra loro in un'unica forma di rappresentazione che consenta l'applicazione di procedure coerenti per rendere esplicite le singole forme strutturali e per permettere il passaggio da una struttura analitica all'altra (Buzzetti, 1996a: 87-88).

Sicché,

il compito di un'edizione digitale può essere al meglio concepito come un tentativo di far corrispondere il modello computazionale per la rappresentazione e l'elaborazione del testo, da una parte al modello concettuale della costituzione del testo e della critica testuale e dall'altra alle procedure e ai metodi dell'analisi dei testi e della loro interpretazione (Buzzetti e Rehbein, 1998: 36).

Ora, poiché l'editore «deve riportare ad unità e coerenza la pluralità dei documenti e la varietà delle forme di rappresentazione attraverso le quali il testo gli è stato trasmesso» e l'interprete «deve, al contrario, risalire da un'unica e coerente forma di rappresentazione del testo ad una molteplicità di strutture ideative ed interpretative tra loro compatibili», se ne trae che, in entrambi i casi, una singola

rappresentazione unitaria deve essere posta in relazione con una molteplicità di strutture diverse tutte compatibili, nell'un caso riportando la molteplicità all'unità, nell'altro procedendo dall'unità alla molteplicità

e che, in entrambi i casi, il modello di dati «può essere lo stesso» (Buzzetti, 1996a: 88). Un modello di dati fondato, per esempio, su 'espansioni di stringhe' (cfr. Thaller, 1996: 252 ss.) è in grado di «ridurre ad unità coerente una molteplicità di diverse e possibilmente sovrapposte rappresentazioni gerarchiche», rispondendo così alle esigenze della critica testuale, o di «derivare, al contrario, da una singola rappresentazione sequenziale una molteplicità di diverse rappresentazioni strutturali» (Buzzetti, 1996b: 255), rispondendo invece alle esigenze dell'analisi del testo (fig. 3).

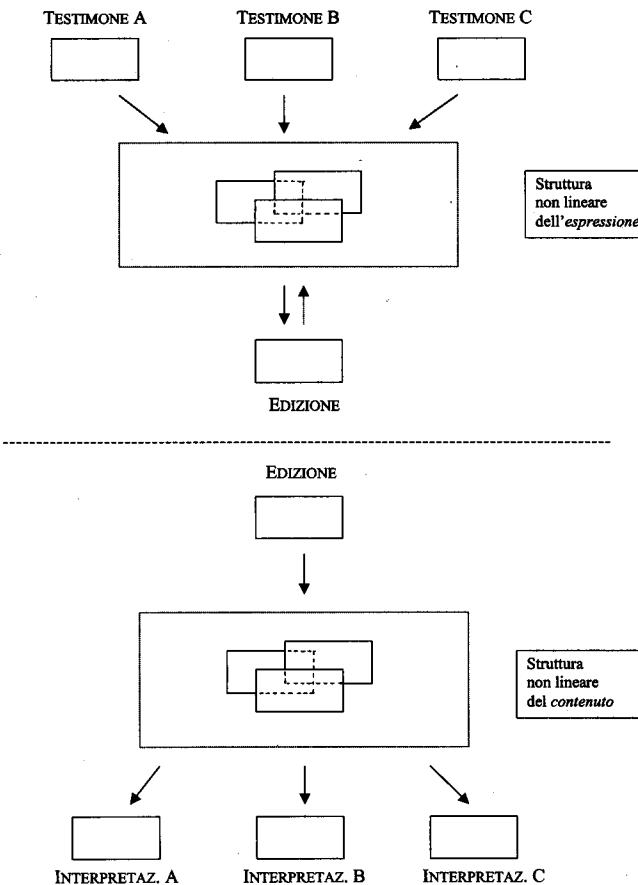


Fig. 3.

Se è vero però che entrambe le operazioni, quella dell'editore e quella dell'interprete, si fondano sostanzialmente sulla stessa relazione uno/molti e possono quindi ricorrere allo stesso modello di dati, resta il fatto che l'una opera sulle strutture dell'espressione (le varianti testuali) e l'altra opera sulla struttura del contenuto (le sue diverse interpretazioni). Entrambe presuppongono un'unica rappresentazione strutturale coerente di natura non lineare, ma nel primo caso essa consiste, per così dire, nella «somma logica» (Thaller, 1993: 64) delle diverse rappresentazioni sequenziali concrete che formano l'intera tradizione testuale, nell'altro, invece, essa costituisce in un certo senso l'«invariante ermeneutico», ossia il «fondamento» (Buzzetti,

1996a: 88) astratto e concettuale «di quella 'affinità'»¹⁵ tra l'ideazione, o le forme di ideazione dell'autore e l'interpretazione, o le forme di interpretazione del lettore, che rende possibile la comprensione del testo» (Buzzetti, 1996a: 88). Tuttavia lo strumento impiegato per applicare algoritmicamente le due pratiche (l'una riguardante l'espressione e l'altra il contenuto) alla rappresentazione digitale del testo è anche in questo caso lo stesso e, di fatto, non è altro che il *markup*. Esso le rende esplicite e le associa alla rappresentazione lineare del testo. Ma anche a questo proposito torna ad essere rilevante, per cercare di definire in modo più preciso questa doppia funzione del *markup*, la distinzione fra l'espressione e il suo contenuto. Il *markup* esprime infatti la relazione fra il modello e l'espressione dei dati.

11. Una precisa idea del rapporto tra l'espressione e il contenuto del testo ci è trasmessa da un'antica massima neoplatonica: «Ogni dialogo è un cosmo e ogni cosmo è un dialogo»¹⁶. Se ne trae che ogni ordine strutturale, ogni modello assegnato al mondo si riflette potenzialmente in un testo che lo rappresenta e che, nello stesso tempo, in ogni testo e in ogni modo di intenderlo e di interpretarlo è contenuta una possibile rappresentazione del mondo, un modo di assegnargli un determinato ordine e una determinata struttura. Sicché ad un contenuto dato corrisponde un'infinita varietà di espressioni e ad un'espressione data corrispondono infiniti modi di intenderne il contenuto. L'indeterminazione del rapporto tra espressione e contenuto è ciò che assicura la dinamicità e la mobilità del testo: alla stessa espressione possono corrispondere contenuti diversi e allo stesso contenuto possono corrispondere espressioni diverse. Se allora è possibile, di conseguenza, «accettare» una «definizione molto ampia» della 'struttura' del testo e la si intende come «l'assieme delle relazioni latenti» tra le sue parti (Segre, 1985: 44), ne viene che anche il numero delle possibili determinazioni strutturali applicabili al testo è potenzialmente infinito. E se il testo (o meglio la sua realizzazione attraverso una forma concreta di espressione) può essere concepito come la linearizzazione di contenuti complessi, il *markup* può essere considerato come lo strumento più idoneo per esplici-

15 Sulla nozione di 'affinità' cfr. Yorck von Wartenburg (1923: 191), citato da Gadamer (1969: 57).

16 Cfr. Westerink (1962: cc. 15 ss.). Sul tema v. anche Coulter (1976: 101-103) e Brisson (1987: 121-128), citati nelle *Notes complémentaires* all'edizione Budé del testo, curata da Westerink *et al.* (1990).

tarne le potenziali relazioni strutturali. Di qui l'importanza di una piena comprensione dello status logico-linguistico del *markup* in relazione all'espressione lineare del testo.

Il *markup* è «inserito e al tempo stesso separabile» dal testo, «fa parte del testo e tuttavia se ne distingue» (Raymond *et al.*, 1992: 3). Esso quindi è *rappresentazione* della struttura e, al tempo stesso, è direttamente *struttura*. Infatti da un lato può essere descritto come una «tecnica per rappresentare la struttura» (Raymond *et al.*, 1992: Abstract) e dall'altro come «struttura» appartenente al testo. È rappresentazione della struttura in quanto è «la denotazione di posizioni specifiche» che identificano stringhe particolari e determinate nella sequenza lineare dei caratteri; ed è esso stesso struttura in quanto la denotazione della posizione è realizzata da «alcuni caratteri specificati (*assigned tokens*)» inseriti e ricompresi nella sequenza di segni che costituisce l'espressione del testo (Raymond *et al.*, 1992: 4). In quanto è *rappresentazione* della struttura, il *markup* è metalinguistico, in quanto è esso stesso *struttura* ed evidenzia tratti strutturali impliciti del testo, ne costituisce un'estensione e ne amplia le risorse espressive. Il *markup* dunque *denota* ed è esso stesso struttura. Per rendersene conto basta considerare una comunissima forma di *markup* come la punteggiatura (cfr. Coombs *et al.*, 1987: 935; Raymond *et al.*, 1992: 9) o, se si vuole, l'uso e la funzione di qualunque segno diacritico. Questa doppia valenza espressiva caratterizza indistintamente tutti i tipi di *markup*, sicché tanto il *markup* vincolato, quanto il *markup* non vincolato possono *denotare* od *essere*, rispettivamente, la struttura dell'espressione o la struttura del contenuto del testo.

Una descrizione adeguata dello status logico del *markup* e della natura ancipite della sua funzione espressiva richiede alcune sommarie indicazioni tecniche. Con l'espressione diretta di aspetti strutturali del testo, il *markup* amplia, per così dire, le risorse espressive del linguaggio-oggetto, introducendo un tipo di espressione che comporta forme di predicazione del secondo ordine, ossia forme di predicazione che si riferiscono non a ciò che il linguaggio rappresenta, ma al suo modo di predicare ordinario o del primo ordine. Un'espressione del linguaggio-oggetto del secondo ordine è quindi un'espressione autoreferenziale, che si riferisce a proprietà astratte o strutturali del linguaggio in quanto forma di rappresentazione, e quindi al suo modo di rappresentare, sia attraverso la forma dell'espressione, sia at-

traverso la forma del contenuto. Non è possibile fornire in questa sede una dimostrazione rigorosa dell'equivalenza tra espressioni del linguaggio-oggetto del secondo ordine ed espressioni metalinguistiche sul significato di espressioni del linguaggio-oggetto del primo ordine ad essa corrispondenti. Basti solo osservare, al riguardo, che l'uso metalinguistico di un certo marcatore descrive la funzione strutturale svolta dallo stesso marcatore usato come espressione diretta della struttura del contenuto o dell'espressione del testo.

12. L'indissolubilità di espressione e contenuto si esprime dunque ancora una volta nel *markup* e ne carica l'uso di implicazioni di ampia portata. Il *markup* è ad un tempo *rappresentazione* e *rappresentazione della rappresentazione*. Il *markup* rappresenta anche il *modo di rappresentare* del testo. Esso esprime ed esplicita una dimensione autoriflessiva del testo. Inteso come *rappresentazione*, il *markup* rappresenta il contenuto o l'espressione del testo, esprimendone direttamente la struttura; inteso come *rappresentazione della rappresentazione*, rappresenta metalinguisticamente la struttura dell'espressione o del contenuto del testo. In ogni caso, si tratta della rappresentazione di un aspetto strutturale del testo, del contenuto o dell'espressione, in cui può esercitarsi una funzione autoriflessiva del rappresentare. Infatti il testo è esso stesso rappresentazione e il *markup* rappresenta, in forme diverse, il modo di rappresentare di questa forma di rappresentazione: o metalinguisticamente, o estendendo le forme espressive del linguaggio-oggetto. La rappresentazione digitale del testo può così trarre indubbio vantaggio proprio dall'uso del *markup*, che comporta l'esplicitazione completa di queste funzioni plurime del linguaggio. Nella rappresentazione digitale del testo, il *markup* diviene il luogo specifico della manifestazione del rapporto tra l'espressione e il contenuto, un rapporto che può essere descritto all'esterno, o reso esplicito all'interno del testo con un ampliamento diretto delle sue risorse espressive.

Così, è proprio attraverso il *markup* che può essere espressa una delle dimensioni più feconde e paradossali del testo e del «testo poetico» in particolare. Procedendo nella lettura di un componimento poetico è consueto il prodursi di improvvisi «cambiamenti di forma» generati dalla tensione costante di diverse strutture concettuali, come l'affacciarsi e il confliggere improvviso di «una nuova e inattesa grammatica» o di un diverso «ordine

ideale». Sicché in un testo poetico si verificano slittamenti continui e ripetuti tra «diversi ordini intratestuali di relazioni testuali»¹⁷, dove le «caratteristiche» percettive, «sonore e visive», insieme con le «caratteristiche semantiche, sintattiche e retoriche» costituiscono altrettanti «campi» di «azione testuale» (McGann, 1999: 81) che interagiscono tra loro e concorrono a rendere il testo essenzialmente un luogo di «instabilità». Il testo poetico può allora, in sostanza, essere visto come un dispositivo, o una «macchina» (McGann, 1999: Draft, 12-15) fatta di segni materiali e concreti, che ne assicura permanentemente la mobilità concettuale. Di conseguenza, il testo, o ciascuno dei suoi elementi, non possono mai essere concepiti come «uguali a se stessi» (McGann, 1999: 81; Draft, *passim*). Incapace di rimanere stabilmente identico a se stesso il testo può essere pensato come un «algoritmo» la cui «logica» riesce ad «essere espressa soltanto con una formula paradossale», come la seguente:

$$A = A \text{ se e solo se } A \neq A \text{ (McGann, 1999: 80-81).}$$

Come di un'«illustrazione» possiamo dire, con Wittgenstein, che «la vediamo come l'*interpretiamo*» (Wittgenstein, 1967, II.xi: 256), allo stesso modo potremmo dire di un testo che lo *comprendiamo* come lo *interpretiamo*. Il celebre esempio di «reversibilità» percettiva (cfr. Galimberti, 1992) citato da Wittgenstein, l'immagine che «si può vedere tanto come una testa di lepre quanto come una testa d'anatra» (Wittgenstein, 1967, II.xi: 256), può essere metaforicamente riferito al testo¹⁸ in quanto sede di fenomeni di reversibilità interpretativa indotti dal «gioco reciproco» dei diversi campi di azione testuale (McGann, 1999: 81). L'instabilità dinamica della «struttura» che «concorre a definirne l'essenza e il significato», caricandosi di volta in volta di «un significato inatteso o di un effetto imprevisto» (Caglioti, 1995: 45), porta il testo ad essere omologabile, nel «campo» (McGann, 1999: 81) della «percezione» stessa «che prelude alla formazione delle idee e allo sviluppo del pensiero» (Caglioti, 1994: 37) a quelle «forme-non forme» (cfr. Thom, 1972) che appartengono «alle classe delle strutture ambigue» naturali

¹⁷ McGann, 1999, Draft: 4-13 (queste ed altre citazioni sono tratte da una stesura più ampia, qui denominata 'Draft', concessa personalmente dall'autore). Cfr. inoltre, su questo tema, il contributo di Ivano Dionigi al presente volume.

¹⁸ Anche un componimento poetico può essere visto come «un'opera lepre-anatra» (McGann, 1999, Draft: 15).

ed artistiche (Caglioti, 1994: 37), la cui caratteristica può «essere definita in termini di coesistenza, in un punto critico, di due aspetti o schemi della realtà che si escludono mutuamente» (Caglioti, 1994: 34).

Ora, tali zone critiche possono essere portate direttamente all'evidenza nel testo dalla «complessa trama» (McGann, 1999: 7) di relazioni testuali esplicitata dal *markup*. La latenza delle relazioni testuali (cfr. Segre, 1985: 44.), la *virtualità* della struttura, si esprime nell'indeterminazione del rapporto tra espressione e contenuto e nella sua legge duale di compensazione. Nel processo interpretativo, l'indeterminazione del contenuto è compensata dalla determinazione dell'espressione; nel processo redazionale e trasmisivo del testo, invece, l'indeterminazione dell'espressione è compensata dalla determinazione del contenuto. Così gli 'scarti (*shifts*)' (cfr. McGann, 1999: 13) concettuali dell'interpretazione sono compensati dalla stabilità dell'edizione di riferimento e la 'fluidità' (cfr. Buzzetti, 1996a; Buzzetti e Rehbein, 1998) della tradizione testuale viene compensata dalla presunta fissità del contenuto, quasi fosse concepito una volta per tutte nell'originario atto ideativo dell'autore. L'instabilità del rapporto tra 'mutazione' (cfr. McGann, 1999: 12) ed 'invarianza' (cfr. Segre, 1985: 29) di espressione e contenuto determina la *mobilità* del testo, che altro non è se non il riflesso dell'indeterminazione della struttura. Il *markup* diviene così il luogo della manifestazione della *virtualità* della struttura e si fa forma specifica dell'esplicitazione dell'Implicito nella rappresentazione digitale del testo. Il *markup* decanta e registra l'instabilità dei fenomeni testuali, riflette visibilmente la storia della trasmissione e dell'interpretazione del testo, ne rispecchia lo sviluppo espressivo e la vitalità concettuale. L'intero apparato critico dell'edizione digitale viene così a dipendere dalla corretta utilizzazione del *markup*.

In quanto espressione della compensazione tra identità e varianza, dell'espressione e del contenuto, rispetto all'indeterminazione della struttura del testo, il *markup* si presta ad essere mezzo di rappresentazione della non-identità del testo con se stesso. Come si è visto, i marcatori possono essere impiegati o come descrizione metalinguistica del testo, o come estensione autoreferenziale del suo linguaggio-oggetto. È questo secondo tipo di impiego che ne fa uno strumento idoneo a rappresentare l'osmosi tra fluidità testuale e fluidità ermeneutica, tra varianza delle lezioni e varianza

delle interpretazioni del testo. La doppia natura, metalinguistica e non, dei marcatori può essere posta in relazione con la convertibilità di espressione e contenuto a quel particolare livello di rappresentazione linguistica che è proprio delle espressioni autoreferenziali. Assunti come espressioni autoreferenziali del secondo ordine in linguaggio-oggetto, i marcatori usati per esprimere diverse strutture concettuali e interpretative del testo possono essere assimilati a distinte varianti testuali. Reciprocamente, diverse varianti testuali possono essere assimilate all'espressione di distinte interpretazioni del testo. In sostanza, il *markup* può trasformare le varianti interpretative in varianti testuali e le variante testuali in varianti interpretative. Come la relazione uno/molti tra identità dell'espressione e varianza del contenuto si può convertire nella relazione uno/molti tra varianza dell'espressione e identità del contenuto, allo stesso modo il *markup* può essere considerato espressione variante di un contenuto sempre identico a se stesso, o esplicazione della varianza di contenuto di un'unica e sempre identica espressione del testo. Il *markup* diventa così uno strumento per trasformare la varianza implicita dell'interpretazione di un'espressione identica nella fluidità esplicita dell'espressione di un identico contenuto. Porzioni testuali variamente codificate, generate da interpretazioni diverse di una stessa espressione, possono essere considerate espressioni sinonime di uno stesso contenuto. È questa caratteristica, legata all'uso autoreferenziale dei marcatori, che permette di rappresentare attraverso il *markup* la non-identità del testo con se stesso. In modo più tecnico, si può dire che il *markup* può essere usato per rappresentare un 'endomorfismo'¹⁹ operante su un insieme di unità testuali distinte o di elementi determinati del testo. In ultima analisi, l'autoriflessività del *markup* costituisce la condizione per poter esprimere in forma esplicita la non-identità del testo con se stesso. In formula,

$$(A = A \text{ se e solo se } A \neq A) \Leftrightarrow A \xrightarrow{\text{markup}} A .$$

Esprimendo la compensazione tra variabilità e invarianza, il *markup* permette di esplorare lo «spazio negativo della testualità» (McGann, 1999, Draft: 16) generato dall'indeterminazione della struttura. La stessa relazione uno/molti può esprimere, in entrambi i sensi, la correlazione compensativa della struttura dell'espressione con la struttura del contenuto e può essere

¹⁹ Cfr., per la nozione di 'endomorfismo', Lawvere e Schanuel (1994: 18).

presa a fondamento dell'adeguatezza funzionale di possibili modelli operativi del testo.

L'esigenza di «trattare analiticamente» (McGann, 1999) l'instabilità del testo vede così nel ricorso a forme adeguate di *markup* uno degli strumenti della sua realizzazione. I nuovi «dispositivi di ordinamento della pagina» introdotti dalla «rivoluzione scribale» del secolo XII svolgevano la stessa funzione riferita alla scrittura, la forma tradizionale di rappresentazione del testo. Attraverso la nuova «architettura visiva», il testo trascritto si piegava «all'immagine mentale della sua struttura» e con «la volontà di usare l'articolazione visiva come mezzo di interpretazione» nasceva «il testo in quanto oggetto» fatto in forma di libro, quella «materializzazione» visiva «dell'astrazione» mentale che ora le tecniche meccaniche introdotte intorno al 1460 «reificano sotto forma di stampato» (Illich, 1994: capp. 6 e 7). L'apparato critico delle edizioni a stampa svolge un'analogia funzione di esplicitazione e di visualizzazione del complesso di relazioni che collega tra loro i diversi testimoni della tradizione testuale. Allo stesso modo, nella rappresentazione digitale, il *markup* può essere utilizzato per l'esplicitazione e la visualizzazione delle relazioni testuali implicite e per rappresentare diversi fenomeni di mobilità testuale. L'uso corretto del *markup* e l'adeguatezza della rappresentazione digitale presuppongono però, come si è visto, il ricorso ad un idoneo *modello del testo*. Sicché la formalizzazione delle procedure critiche ed interpretative e la specificazione del modello operativo restano l'imperativo essenziale per la produzione di edizioni digitali in grado di garantire l'esaustività e la funzionalità di questa nuova e più complessa forma di rappresentazione del testo.

BIBLIOGRAFIA

- BARNARD D., HAYTER R., KARABABA M., LOGAN G., MCFADDEN J., 1988. *SGML-Based markup for literary texts: Two problems and some solutions*. Computers and the Humanities, 12: 265-276.
- BATESON G., 1976. *Forma, sostanza e differenza*. In: *Verso un'ecologia della mente*. Adelphi, Milano: 464-484.

- BRISSON L., 1987. *Le discours comme univers et l'univers comme discours: Platon et ses interprètes néo-platoniciens*. In: *Le texte et ses représentations*. Presses de l'École normale supérieure, Paris: 121-128.
- BURNARD L., 1991. *An introduction to the text encoding initiative*. In: D. GREENSTEIN (ed.), *Modelling historical data*. Max-Plank-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, St. Katharinen: 81-91.
- BURNARD L., SPERBERG-MCQUEEN C. M., 1991. *Living with the guidelines: an introduction to TEI tagging*. Text Encoding Initiative, Document Number: TEI EDW18, March 13.
- BUZZETTI D., 1995. *Image processing and the study of manuscript textual traditions*. Historical Methods, 28: 145-154.
- BUZZETTI D., 1996a. *Il testo 'fluido'. Sull'uso dell'informatica nella critica e nell'analisi del testo*. In: L. FLORIDI (a cura di), *Filosofia & informatica*. Atti del primo incontro italiano sulle applicazioni informatiche e multimediali nelle discipline filosofiche (Convegno Nazionale della Società Filosofica Italiana: Roma, 23-24 novembre 1995). Paravia, Torino: 85-93.
- BUZZETTI D., 1996b. *Digital editions: variant readings and interpretations*. In: *ALLC-ACH'96 Conference Abstracts*. University of Bergen: 254-256.
- BUZZETTI D., 1999. *Text representation and textual models*. In: *ACH-ALLC'99 Conference Proceedings*. University of Virginia, Charlottesville VA: 219-222.
- BUZZETTI D., REHBEIN M., 1998. *Textual fluidity and digital editions*. In: M. DOBREVA (ed.), *Text variety in the witnesses of medieval texts*. Proceedings of the International Workshop (Sofia, 21-23 September 1997). Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Sofia: 14-39.
- BUZZETTI D., TABARRONI A., 1991. *Informatica e critica del testo. Il caso di una tradizione 'fluida'*. Schede umanistiche, n.s., I, n. 2: 185-193.
- CAGLIOTI G., 1994. *Simmetrie infrante nella scienza e nell'arte*. 2a ed., Città Studi Edizioni, Milano.
- CAGLIOTI G., 1995. *Eidos e psiche. Struttura della materia e dinamica dell'immagine*. Ilisso, Nuoro.
- COOMBS J. H., RENEAR A. H., DE ROSE S. J., 1987. *Markup systems and the future of scholarly text processing*. Communications of the ACM, 30: 933-947.
- COULTER J., 1976. *The literary microcosm: theories of interpretation of the later neoplatonists*. Columbia Studies in the Classical Tradition, II. Brill, Leiden.
- COVER R., DUNCAN N., BARNARD D. T., 1981. *The progress of SGML (Standard Generalized Markup Language): extracts from a comprehensive bibliography*. Literary and Linguistic Computing, 6: 197-209.
- DAY A. C., 1984. *Text processing*. Cambridge University Press, Cambridge.

- DE ROSE S. J., DURAND D. D., MYLONAS E., RENEAR A. H., 1990. *What is text, really?* *Journal of Computing in Higher Education*, 1(2): 3-26.
- GADAMER H. G., 1969. *Il problema della coscienza storica* (1963). Trad. it. di G. Bartolomei. Guida, Napoli.
- GALIMBERTI G., 1992. *Reversibilità*. In: *Dizionario di psicologia*. UTET, Torino: 812-813.
- GOLDSCHLAGER L., LISTER A., 1988. *Computer science: a modern introduction*. 2nd ed., Prentice Hall, London.
- HERSTEIN I. N., 1975. *Topics in Algebra*. 2nd ed., John Wiley, New York.
- HJELMSLEV L., 1968. *I fondamenti della teoria del linguaggio* (1943, 1961²). Trad. it., Einaudi, Torino.
- ILLICH I., 1994. *Nella vigna del testo. Per una etologia della lettura* (1993). Trad. it. di A. Serra e D. Barbone. Cortina, Milano.
- JOLOBOFF V., 1989. *Document representations: concepts and models*. In: J. ANDRÉ, R. FURUTA, V. QUINT (eds.), *Structured documents*. Cambridge University Press, Cambridge: 75-105.
- KIRSCHENBAUM M. G. et al., 1999. *Refining our notions of what (digital) images really are*. In: *ACH-ALLC'99 Conference Proceedings*. University of Virginia, Charlottesville VA: 55-58.
- LA TORRE M., 1994. *Principi di informatica*. La Nuova Italia, Firenze.
- LAWVERE F. W., SCHANUEL S. H., 1994. *Teoria delle categorie. Un'introduzione alla matematica*. Trad. it. di M. Mondolfo. Muzzio, Padova.
- LÉVY P., 1990. *Les technologies de l'intelligence: l'avenir de la pensée à l'ère informatique*. La Découverte, Paris.
- MCGANN J. J., 1997. *The rationale of hypertext*. In: K. SUTHERLAND, *Electronic text: investigations in method and theory*. Clarendon Press, Oxford: 19-46.
- MCGANN J. J., 1999. *What is text?* In: *ACH-ALLC'99 Conference Proceedings*. University of Virginia, Charlottesville VA: 80-81.
- MIROSLAVJEVIC M., VITALI F., WATTERS C., 1999. *Introduction*. In: *HTF/VD1: Workshop on Virtual Documents, Hypertext Functionality and the Web*, organized by M. Miroslavjevic, F. Vitali, C. Watters, at *WWW8: The Eighth International WorldWideWeb Conference* (Tuesday, May 11, 1999). Toronto, Canada: <URL: <http://www.cs.unibo.it/~fabio/VD99/intro.html>>.
- NAESS A., 1953. *Interpretation and preciseness: a contribution to the theory of communication*. I kommisjon hos J. Dybwad, Oslo.
- NELSON TH. H., 1997. *Embedded markup considered harmful*. In: *XML: principles, tools, and techniques* («The World Wide Web Journal», 2, 1997, no. 4). O'Reilly, Cambridge MA: 129-134.

- RAMNEZ S., CRAMPES M., 1999. *Conceptual documents and hypertext documents are two different forms of virtual document*. In: *HTF/VD1: Workshop on Virtual Documents, Hypertext Functionality and the Web*, organized by M. Miroslavjevic, F. Vitali, C. Watters, at *WWW8: The Eighth International WorldWideWeb Conference* (Tuesday, May 11, 1999). Toronto, Canada: <URL: <http://www.cs.unibo.it/~fabio/VD99/ranwez/ranwez.html>>
- RAYMOND D. R., TOMPA F. W., WOOD D., 1992. *Markup reconsidered*. In: *First International Workshop on Principles of Document Processing* (October 22-23). Washington D.C.: [1-20] <URL:<http://www.csd.uwo.ca/staff/drraymon/.papers/markup-ps>>
- RAYMOND D. R., TOMPA F., WOOD D., 1995. *From data representation to data model: Meta-semantic issues in the evolution of SGML*. Computer Standards and Interfaces, 10: [1-22] <URL:<http://www.csd.uwo.ca/staff/drraymon/.papers/sgml-ps>>
- RENEAR A., MYLONAS E., DURAND D., 1996. *Refining our notion of what text really is: The problem of overlapping hierarchies*. In: N. IDE (ed.), *Research in humanities computing*, vol. 4. Selected papers from the ALLC/ACH Conference (Christ Church, Oxford, April 1992). Clarendon Press, Oxford: 263-280.
- SEGRE C., 1985. *Avviamento all'analisi del testo letterario*. Einaudi, Torino.
- SMITH J. M., 1992. *SGML and related standards: document description and processing languages*. Ellis Horwood, New York.
- SPERBERG-MCQUEEN C. M., BURNARD L. (eds.), 1994. *Guidelines for text encoding and interchange*. Text Encoding Initiative, Chicago - Oxford.
- SPERBERG-MCQUEEN C. M., HUITFELDT C., 1999. *GODDAG: a data structure for overlapping hierarchies*. In: *ACH-ALLC'99 Conference Proceedings*. University of Virginia, Charlottesville VA: 198-201.
- STEER G., 1993. *Textkritik und Textgeschichte. Editorische Präsentation von Textprozessen: Das Nibelungenlied. Der Schwabenspiegel. Die Predigten Taulers*. In: *Methoden und Probleme der Edition mittelalterlicher deutscher Texte*. Hg. von Rolf Bergmann und Kurt Gärtner (Beihefte zu *Editio*, 4). Niemeyer, Tübingen.
- THALLER M. (ed.), 1991. *Halbgraue Reihe zur Historischen Fachinformatik*. (Serie C: *Datenbasen als Editionen*). Max-Planck-Institut für Geschichte. Scripta Mercaturae Verlag, St. Katharinen.
- THALLER M., 1993. *Historical information science: is there such a thing? New comments on an old idea*. In: T. ORLANDI (ed.), *Discipline umanistiche e informatica. Il problema dell'integrazione*. Seminario (Roma, 8 ottobre 1991). Contributi del Centro Linceo Interdisciplinare B. Segre, 87. Accademia Nazionale dei Lincei, Roma: 51-86.
- THALLER M., 1996. *Text as a data type*. In: *ALLC-ACH'96 Conference Abstracts*, University of Bergen: 252-254.

- THALLER M., 1999. *A proposal for a humanities text processing protocol*. In: *ACH-ALLC'99 Conference Proceedings*. University of Virginia, Charlottesville VA: 134-135.
- THOM R., 1972. *Stabilité structurale et morphogenèse*. Benjamin, Reading Mass.
- WATTERS C., SHEPHERD M., 1999. *Research issues for virtual documents*. In: *HTF/VD1: Workshop on Virtual Documents, Hypertext Functionality and the Web*, organized by M. Miroslavjevic, F. Vitali, C. Watters, at *WWW8: The Eighth International WorldWideWeb Conference* (Tuesday, May 11, 1999). Toronto, Canada: <URL: <http://www.cs.unibo.it/~fabio/VD99/shepherd/shepherd.html>>.
- WESTERINK L. G. (ed.), 1962. *Anonymous prolegomena to Platonic philosophy*. North-Holland, Amsterdam.
- WESTERINK L. G., TROUILLARD J., SÉGONDS A. PH. (eds.), 1990. *Prolégomènes à la philosophie de Platon*. Les Belles Lettres, Paris.
- WITTGENSTEIN L., 1967. *Ricerche filosofiche* (1953). Ed. it. a cura di M. Trinchero. Einaudi, Torino.
- YORCK VON WARTENBURG P., 1923. In: *Briefwechsel zwischen Wilhelm Dilthey und dem Grafen Paul Yorck von Wartenburg, 1877-1897*. Niemeyer, Halle (Saale).