

XML e i testi contemporanei: problemi di codifica, rappresentazione e analisi

1

Domenico Fiormonte (Università Roma Tre [fiormont@uniroma3.it]) **Lorenzo Geri** (Università di Roma “La Sapienza” [lorenzogeri@yahoo.com]) **Daniele Silvi** (Università di Roma Tor Vergata, [silvi@lettere.uniroma2.it])

1. Fruizione e rappresentazione del testo digitale

In questo intervento presentiamo uno studio sulla codifica e l'analisi digitale di alcuni dattiloscritti dello scrittore italiano Vincenzo Cerami. Tale materiale fa parte di un ampio corpus messo a disposizione dell'autore nell'ormai lontano 1994 per una ricerca sull'influsso del computer sulla scrittura (Fiormonte 1995). Il tempo che ci separa da quei primi esperimenti ci permette, oltre che di operare un bilancio, di esporre alcune riflessioni su come sia cambiata la concezione del testo digitale negli ultimi dieci anni. Non sembri paradossale che in un articolo che propone soluzioni secondo le linee guida della TEI (*Text Encoding Initiative*) si esprimano perplessità su talune questioni che riguardano proprio XML-TEI e in particolare la rappresentazione e la visualizzazione del processo di scrittura – oggetto principale di questa ricerca. Poiché in larga misura gli strumenti di digitalizzazione che di volta in volta impieghiamo *costruiscono l'accesso al documento*, è evidente che non è auspicabile né utile pretendere di elaborare una teoria e una pratica del testo digitale indipendenti dal tempo, cioè valide una volta per tutte. Quanto esporremo in questa introduzione dunque ha lo scopo di stimolare gli addetti ai lavori, e noi stessi, a riflettere e possibilmente ad andare oltre le attuali tecnologie.

Se fino a metà degli anni Novanta la preoccupazione principale dell'editore è stata quella di *conservare* nel modo più fedele possibile l'informazione contenuta nella fonte originale, negli ultimi dieci anni, soprattutto sotto la pressione delle forme ‘native’ della comunicazione digitale, l'attenzione si è spostata sulla *fruizione* del testo. A bene vedere la ‘priorità conservativa’ è stato il sintomo di una tendenza più generale dell'informatica. Nell'infanzia della digitalizzazione infatti gli sforzi della comunità scientifica si sono concentrati sulla realizzazione di strumenti per il controllo e la gestione dei dati, finendo per disegnare confini tutto sommato angusti per il testo – che non è solo un oggetto informativo, ma anche

comunicativo e semiotico. Quest'ultima dimensione del testo, che possiamo definire a un tempo "materiale" e "performativa", è rimasta abbastanza inesplorata nel campo della filologia digitale, nonostante i ripetuti appelli degli eredi della *textual bibliography* alla "primacy of the physical object" (McGann 2005: 239) e le aperture alla pragmatica, cioè al tessuto di relazioni che lega il testo al suo uso (Shillingsburg 2006). Se è vero che la codifica rafforza e in certi casi amplia le possibilità di analisi critica del testo (Albonico 2005: 256) è altrettanto vero che, mentre si sforza di rappresentare tutti gli aspetti del documento originale (nato in un contesto mediale diverso), finisce per trascurare le caratteristiche e le possibilità del medium all'interno del quale si muove. Questo problema ha origini e ragioni complesse che è impossibile riassumere in poche battute. Tuttavia due cause possono essere additate: 1) da un punto di vista generale è indubbio che sulle attuali tecnologie pesi ancora l'ipoteca del paradigma di lettura e scrittura della stampa. Tale paradigma ha guidato anche la nascita di SGML, antenato di XML; 2) l'esigenza di ricostruzione del testo ereditata dalle filologie classica e medievale si è sposata felicemente con l'idea di modello astratto (Buzzetti 2002), e con la promessa che a partire da un unico "meta-testo" (il documento codificato) fosse possibile ricavare diverse edizioni su diversi supporti. Una delle caratteristiche dei linguaggi di markup che è sembrata meglio rispondere alle esigenze dell'ecdotica è infatti quella di separare la struttura dei contenuti dalla loro presentazione (Orlandi 1999: 92; Albonico 2005: 245). Ma questa separazione fra codifica di un documento e sua rappresentazione finale (a partire da un file XML possibile produrre diversi tipi di output, da una pagina web a un file PDF) finisce per separare – praticamente e ontologicamente – il piano del medium da quello del contenuto:

(...) the belief that a text can be rearranged and moulded by technical means, presupposes that content and presentation can be treated independently, as logically different features: it rests on the false supposition that any kind of written or verbal content can be represented at will in any medium and for whatever purpose (Hillesund 2005).

Le capacità insite nei linguaggi di markup di astrarre dal contenuto, chiave del loro successo (portabilità, accessibilità, duttilità, ecc.), coincidono dunque con il loro limite: poiché il grado di fedeltà alla fonte originaria è proporzionale al grado di astrazione che il linguaggio rende possibile, più aumenta la complessità del documento, più complesso e verboso il markup e maggiori le difficoltà di visualizzazione e dunque di accesso a esso.

In conclusione, se non è certo possibile condannare l'esigenza conservativa, è pure vero che l'opera di creazione della matrice meta-testuale non deve prevalere sulle esigenze del lettore finale. Per questo oggi sembra abbastanza chiaro che per il rilancio del contenuto digitale

abbiamo bisogno di interfacce e che, riguardo al testo letterario, “what was one a problem of page layout and typography has become a problem of human-computer interaction design” (Schmidt 2006b: 8). Questo specifico problema riguarda anche i nostri testi, per i quali abbiamo realizzato una visualizzazione usando i fogli di stile XSLT (cfr. § 4) che, pur nella loro coerenza rappresentazionale, sono ancora lontani dall’obiettivo di una piena ed efficace fruizione. Questa insoddisfazione non è il segno dell’inadeguatezza degli strumenti – giacché ricadremmo nella trappola di chiedere al digitale di risolvere i problemi del mondo della carta – ma è il frutto di un cammino decennale di ricerca che ci ha portato, dall’idea di archivio-risorsa per studiosi, a quella di strumento d’interazione e dialogo con gli utenti. Accostandoci al problema della rappresentazione/fruizione del processo di scrittura ci siamo resi conto, ad esempio, che le forme ‘native’ della scrittura digitale ponevano delle domande del tutto nuove ai filologi (Wardrip-Fruin 2005), illuminando e talvolta scuotendo il lavoro di edizione dei testi. Le tecnologie di produzione *non sono più disgiunte da quelle di conservazione*, come la difficile convivenza stampa-digitale ci induce a credere. E il riflesso delle forme native della comunicazione digitale si riverbera all’indietro, colpendo al cuore il nostro sistema di valori: modi, sguardi, metodologie.

2 L’esperienza di *Digital Variants*

2.1 I progetti di ricerca in Europa e negli Stati Uniti dedicati alla codifica delle fonti primarie (*primary sources*) dimostrano la vitalità di questo campo di studi, e allo stesso tempo come la questione si possa affrontare secondo una grande varietà di approcci.² Tale varietà è dovuta anche al fatto che il *tagset* della TEI P4 dedicato alla codifica dei manoscritti (*Transcription of Primary Sources*) non si adatta alla codifica di tutte le tipologie di fonti primarie, in quanto è stato ideato per un’edizione critica in forma elettronica (di fatto si tratta di una codifica XML di un’edizione critica cartacea):

There may be several reasons for encoding manuscripts, and the end result may not necessarily be the production of a critical edition. [...] it would be semantically incorrect to use the Transcription of Primary Sources subset to encode several layers of additions, deletions and variation in the transcription of one document³.

Per quanto riguarda la situazione italiana, la codifica dei testi letterari è stata per lo più realizzata sulle opere canoniche della nostra ricca tradizione letteraria (sino al primo Novecento). Ne è conseguito un modello di edizione elettronica pensato per la costituzione di banche date testuali, mentre gli esperimenti di codifica ‘filologica’ si sono concentrati sui

problemi, ingenti, che riguardano la formalizzazione dei manoscritti. In Italia la codifica degli avantesti è un campo quasi del tutto inesplorato.⁴

2.2 La possibilità di diversi approcci alla codifica dei testi contemporanei è stata sperimentata nell'ambito del progetto *Digital Variants*⁵. Nel codificare gli avantesti di autori contemporanei il gruppo di ricerca si è trovato di fronte a del materiale eterogeneo: da una parte la prima redazione manoscritta del libro di poesie *Ora serrata retinae* di Valerio Magrelli; dall'altra i dattiloscritti con correzioni autografe di racconti di Vincenzo Cerami. Nel primo caso, per rendere conto di una situazione testuale fluida, stratificata e complessa è nata la *Magrelli machine*;⁶ nel secondo caso è stata realizzata una codifica in HTML del testo definitivo (quello in volume) e di tre stesure precedenti. I file HTML sono stati visualizzati in *frames* per analizzare le differenze, utilizzando *java-script*. Il risultato è la *Vincenzo Cerami Variants Machine* realizzata da Barbara Lotti, Denise Rana e Mario Macciocca⁷, ai quali si è aggiunta la cura linguistica di Cinzia Pusceddu, dell'Università di Edimburgo. La finalità di tale strumento è essenzialmente didattica, dal momento che permette ai docenti di lingua italiana all'estero di mostrare agli studenti in maniera piuttosto semplice e intuitiva il processo genetico e creativo di un testo letterario, commentando le scelte linguistiche.

3 Codifica dei dattiloscritti

3.1. Scopo del progetto dedicato a Vincenzo Cerami è quello di studiare il processo editoriale e compositivo attraverso la codifica e l'analisi informatica degli avantesti di una raccolta di racconti, *La gente* (Torino, Einaudi, 1991). Possiamo seguire passo passo il costituirsi del libro, grazie alle diverse stesure dei racconti che costituiscono il libro che l'autore ci ha messo a disposizione⁸. Ai fini di questo contributo ci limitiamo a elencare le tipologie di redazioni da noi codificate: ogni racconto, tranne delle eccezioni del quale qui non diamo conto, nasce come contributo per la pagina domenicale del quotidiano romano *Il Messaggero*, ed entra a far parte di una rubrica dal nome "L'indiscreto". Le pubblicazioni sul giornale sono le prime testimonianze di cui disponiamo e che abbiamo codificato. In seguito Cerami lavora sui racconti, partendo da una stesura dattiloscritta che riproduce la lezione pubblicata sul giornale; alla prima stesura dattiloscritta con correzioni ne seguono altre tre. Dopo le tre redazioni dattiloscritte Cerami interviene sulle bozze che sono state incluse nella codifica .

3.2 Le caratteristiche del materiale che abbiamo codificato sono le seguenti:

(a) Non possediamo una stesura manoscritta. Questa circostanza comporta che i testi che analizziamo presentino uno stato avanzato di formalizzazione: i racconti hanno già raggiunto una struttura narrativa stabile, destinata a subire pochi cambiamenti da una redazione all'altra. Inoltre le correzioni dell'autore non sono mai profonde, anche se non per questo, come vedremo, tutti i fenomeni correttori possono essere codificati senza trovarsi a dover affrontare il problema delle *overlapping hierarchies*.

(b) Il fatto che le correzioni siano relative ad una stesura avanzata (in quanto posteriori alla pubblicazione sul quotidiano) fa sì che sia possibile confrontare agevolmente le diverse stesure e ricostruire la funzione di ogni singola correzione. Per questo abbiamo scelto di effettuare una codifica delle correzioni che non registri soltanto il fenomeno, ma cerchi, tramite gli attribuiti, di codificarne anche la motivazione.

Nella codifica HTML precedente, come già detto, l'attenzione era incentrata sulla visualizzazione delle cancellazioni e l'aggiunta di parole, mentre i fenomeni correttori di diverso tipo non erano presi in considerazione.

3.3 La nuova codifica XML secondo la TEI P4 è stata pensata per dare conto del fenomeno correttorio in un modo quanto più completo. Abbiamo lavorato su tre livelli:

(a) abbiamo codificato i fenomeni correttori che riguardano direttamente la lezione del testo, utilizzando il tagset standard *Transcription of Primary Sources*;

(b) abbiamo codificato quelle parti del dattiloscritto che sono cerchiare, sottolineate o evidenziate in altro modo dall'autore;

(c) abbiamo inserito nella nostra codifica un'interpretazione di entrambi i fenomeni correttori.

3.3.1 Per quanto riguarda la trascrizione dei processi correttivi che incidono sulla lezione del testo abbiamo utilizzato il codset *Transcription of Primary Sources*. I tag e <add> permettono, nei casi semplici, di codificare anche il tempo e lo spazio del fenomeno correttivo:

Tabella 1.

fenomeno	spazio	tempo (stesure)
----------	--------	-----------------

		hand=""/ rend=""
<add>	place=""	hand=""/ rend=""

L'attributo *place* è utilizzato generalmente per indicare se l'intervento è supralineare o sublineare; nel nostro caso abbiamo sostituito ai valori generalmente in uso l'indicazione del margine destro o sinistro. Naturalmente le coordinate che si possono fornire con un simile sistema non possono che essere approssimative. Per quanto riguarda la codifica delle correzioni, l'attributo *hand* permette di segnalare un cambio di inchiostro, che può indicare una successione nelle stesure. Si può immaginare anche che lo scrittore utilizzi penne diverse contemporaneamente, per indicare un diverso tipo di intervento: ma in casi come questi è difficile che il codificatore possa risalire al codice che lo scrittore ha elaborato autonomamente.

Nel corso della nostra codifica ci siamo imbattuti in una serie di indicazioni correttorie che non sono identificabile né con il fenomeno descritto con né con quello descritto da <add>.

In questo caso abbiamo deciso di utilizzare un tag generico come <seg>, la cui sintassi permette, tramite l'uso degli attributi *rend* e *type* di codificare fenomeni complessi e, soprattutto, non previsti dalla DTD TEI. Per quanto riguarda la descrizione dei fenomeni correttori abbiamo utilizzato l'attributo *rend*, al quale abbiamo associato una descrizione del fenomeno così come si presenta sulla pagina:

Tabella 2.

sottolineatura	<seg rend="sottolineato"></seg>
cerchiatura	<seg rend="tratto cerchiato"></seg>
evidenziazione laterale	<seg rend="tratto a matita verde, margine sinistro"></seg>
trattino	<seg rend='tratto a matita verde, margine sinistro'>-</seg>

Come si nota dalla Tabella 2, dal momento che il tag <seg> non prevede la possibilità di utilizzare l'attributo *place*, abbiamo indicato le informazioni relative allo spazio all'interno

delle descrizioni assegnate all'attributo *rend*. Questa scelta si basa sulla convinzione che talora è necessario 'forzare' la DTD piuttosto che perdere informazioni. Sarebbe ad ogni modo possibile ovviare al problema modificando la DTD. Si potrebbe prevedere per <seg> l'attributo *place*. Ma se si sceglie di modificare la DTD, al nostro modo di vedere, sarebbe più coerente progettare un nuovo tagset, piuttosto che modificare dei tag esistenti.

Nel codificare i dattiloscritti di Cerami ci siamo resi conto dell'utilità di interpretare di volta in volta gli interventi dell'autore, anche in quei casi in cui i segni sulla pagina non sono immediatamente riconducibili a un intervento sul testo (per intenderci Cerami sottolinea, cerchia, evidenzia in vari modo dei punti del testo ma, per il momento, non cancella né aggiunge parole). Naturalmente una codifica del genere è tanto profonda quanto potenzialmente invasiva. Per distinguere la lettura del dattiloscritto dalla nostra interpretazione dei fenomeni correttivi abbiamo ritenuto utile ricorrere all'attributo *type*, associandolo a una serie di azioni correttive:

Tabella 3.

eliminazione di una ripetizione	type="ripetizione"
identificazione di un episodio da rivedere	type="rivedere"
identificazione di un episodio	type="identificazione_di_un_episodio"

La terminologia che abbiamo utilizzato talora consiste nella descrizione del motivo per il quale l'autore ha agito, ad esempio "ripetizione", che indica lo sforzo di evitare una ripetizione, oppure "identificazione di un episodio; in altri casi la terminologia riporta degli imperativi che lo scrittore presumibilmente rivolge a sé stesso, come "rivedere".

Per interpretare quello che l'autore indica con dei segni come la sottolineatura, la cerchiatura delle parole ecc., è necessario spesso confrontare il passo evidenziato con la lezione della stesura successiva. In questo modo si può capire il motivo per il quale l'autore ha sottolineato le parole: nel maggior parte dei casi si tratta di una ripetizione che, nella stesura successiva, viene risolta ricorrendo ad un sinonimo; in altri casi l'autore sottolinea un'intera sequenza narrativa che nella seguente stesura riduce oppure cancella del tutto.

3.3.2 Esempi di codifica

Figura 1.

—Nelle notti d'agosto il centro storico diventa addirittura
terrorizzante. Gli antichi fregi senza luci, le lunghe file dei
negozi spenti, le grandi pareti dei palazzi dalle finestre chiuse
e dai portoni sbarrati acquistano di colpo l'immagine di una
persona amica ^{MA} di colpo stravolta dalla pazzia, pronta a saltarti
addosso e a sbranarti senza pietà.

In questa pagina dapprima Cerami nota la ripetizione del sintagma avverbiale: “di colpo”; quindi decide di sostituire il secondo “di colpo” con l’avverbio “ma”.

Prima o dopo tale operazione l’autore traccia un trattino all’inizio della frase. Studiando tutti i trattini che ricorrono nei dattiloscritti ci siamo resi conto che ogni trattino segnala l’inizio di una sequenza narrativa. Tant’è che in un caso Cerami elimina una porzione di testo che corrisponde esattamente alla sequenza narrativa indicata in tale modo.

Per codificare questa sequenza correttiva secondo i principi esposti in precedenza avremo:

```
<p><seg type='individuazione sequenza narrativa' rend='tratto  
a matita verde, margine sinistro'>-</seg><-</seg>  
  
...  
  
<seg rend="tratto cerchiato a matita  
verde" type="ripetizione">di colpo</seg> l'immagine di una  
persona amica <seg rend='tratto cerchiato a matita  
verde' type='ripetizione' >di colpo<add hand='autore'  
place='supralinear'>ma</add></seg> stravolta dalla pazzia,  
pronta a saltarti addosso e a sbranarti senza pietà.</p>
```

Veniamo ad un altro esempio:

Figura 2.

Domenica sera, dopo aver passato tutta la giornata in casa a finire un lavoro che mi aveva costretto a rimanere ^{nella} in questa città abbandonata ~~nelle~~ ^{alle} grinfie del caldo, uscii a piedi, a caccia di un ristorante. Volevo fare due passi per sgranchirmi la schiena e anche perché avevo voglia di vedere qualche faccia vivente. Sono infatti un appassionato di facce: mi piace guardarle e immaginare da quali ambienti vengono fuori, quali pensieri passano per quelle teste; mi piace figurarmi la casa e la vita delle persone

In questa pagina l'autore ha eliminato il sintagma "in questa" sostituito con "nella"; ha eliminato "nelle" sostituendolo con "alle" (forse per evitare una ripetizione); infine, ed è il fenomeno più complesso, ha evidenziato la frase che va da "Sono infatti" sino a "la vita di quelle persone" tramite due tratti sul margine sinistro. In questo caso la codifica sarà:

```
<p>Domenica sera, dopo aver passato tutta la giornata in casa a finire un lavoro che mi aveva costretto a rimanere <add hand='autore' place='supralinear'>nella</add><del hand='autore'>in questa</del>città abbandonata <add hand='autore' place='supralinear'>alle</add> <del hand='autore'>nelle</del> grinfie del caldo, uscii a piedi, a caccia di un ristorante. Volevo fare due passi per sgranchirmi la schiena e anche perché avevo voglia di vedere qualche faccia vivente. Sono infatti un appassionato di facce: mi piace guardarle e immaginare <seg rend="doppio_tratto, sul margine sinistro" hand="matita nera" type="da_rivedere">da quali ambienti vengono fuori, quali pensieri passano per quelle teste; mi piace figurarmi la casa e la vita delle persone</seg>
```

Come ultimo esempio discuteremo il fenomeno più complesso che abbiamo incontrato:

Figura 3.

colpì con uno schiaffo sonoro. Lei scoppiò a piangere. Lui allora si alzò e la trascinò via con la forza, picchiandola duramente.

~~Stavo per intervenire. Mi alzai, ma non mi mossi.~~ Li vedevo tornare verso il portone da cui erano usciti, ~~e capii che in quei loro rituali io non dovevo entrare più.~~ La mia parte l'avevo involontariamente recitata fino in fondo. Per tutta la cena non ero stato che lo zimbello nelle mani di quei due strani innamorati. M'avevano tirato dentro le loro beghe. E ora ~~me~~ ^M li immaginavo rientrare nella piccola camera accaldata, dove avrebbero continuato il loro complicato gioco d'amore.

Tornando a casa le vie del centro mi apparivano ancora più

In questa pagina Cerami cancella la frase “Stavo per intervenire. Mi alzai, ma non mi mossi”; quindi segna le frasi che vanno da “Li vedevo” sino a “le loro beghe”, per segnalare che l'intero episodio non lo soddisfa. In seguito⁹ l'autore cancella tre frasi, si suppone una di seguito all'altra. Quindi decide di cassare l'intero episodio (e si tratta proprio di quella porzione di testo che aveva segnato inizialmente), e lo fa con una grande X, che anche iconicamente indica l'insoddisfazione per quelle frasi. Per finire inserisce una maiuscola nel punto nel quale ha deciso di far continuare il racconto. Tale situazione andrebbe rappresentata nel modo seguente:

```
<del>Stavo per intervenire. Mi alzai, ma non mi mossi.</del>
<seg type='rivedere' rend='doppio tratto verticale a margine destro'>
<seg type='cancellatura stratificata' rend='croce'>Li vedevo tornare verso
il portone da cui erano usciti <del> e capii che in quei loro rituali io
non dovevo entrare più.</del> La mia parte l'avevo <del> involontariamente
recitata fino in fondo.</del> Per tutta la cena non ero stato che lo
zimbello nelle mani di quei due strani <del> innamorati. M'avevano tirato
dentro le loro beghe. E ora m</del></seg><add>M</add>e li immaginavo
...
```

Come si noterà è soltanto il caso a far sì che il fenomeno descritto dal primo <seg> non interferisca con il fenomeno descritto dal secondo <seg>, dando vita ad un caso di *orvelapping hierarchy*.

3.3.3 Prima di andare avanti possiamo enucleare le considerazioni principali che la nostra codifica ci ha portato a formulare:

(a) sarebbe utile sviluppare un *codset* che si occupi di quei fenomeni correttori che non riguardano immediatamente la lezione del testo; tale codset potrebbe essere pensato come un sottoinsieme del codset *Transcription of Primary Sources*;

(b) l'uso del tag `<seg>` e degli attributi, in particolare di *type*, ci permette di forzare i limiti della DTD TEI senza per questo dover immaginare nuovi tag;

(c) una codifica semantica può essere utilizzata unitamente a una codifica descrittiva e permette di aumentare la complessità della descrizione, senza per questo complicare eccessivamente lo schema di codifica.

4. La visualizzazione con XSLT

Grazie all'uso di XSLT gli attributi possono essere utilizzati anche per potenziare la visualizzazione del file XML. Considerando poi che nella precedente codifica HTML l'attenzione era incentrata solamente sulla visualizzazione di alcuni fenomeni, si è reso necessario un meccanismo più convincente di visualizzazione che sfruttasse appieno la nuova codifica XML, valorizzando il lavoro svolto. Una trasformazione XSLT è sembrata la scelta migliore, grazie alle strutture di indagine all'interno dei tag e degli attributi che questo linguaggio mette a disposizione. La sintassi XSLT è la medesima di XML e ciò rende il lavoro più coerente e maggiormente aperto a eventuali, futuri, ritocchi alla codifica.

Per ora è stato creato un solo foglio di stile XSL che tenesse conto di tutti i fenomeni codificati, ma per lo sviluppo futuro del progetto provvederemo a separare i fogli di stile per ciascuna delle versioni analizzate. Quello che segue è un breve esempio del foglio di stile utilizzato:

```
<xsl:template match="seg">
```

```
<xsl:choose>
```

```

<xsl:when test= "@type='ripetizione'">

    <font color="green" size="4">

        <u>
            <xsl:apply-templates/>
        </u>
    </font>
</xsl:when>

<xsl:when test= "@type='individuazione sequenza
narrativa'">

    <font color="green" size="5">
        <h3><xsl:value-of select="@type"/></h3>
        [ <xsl:apply-templates/> ]
    </font>

</xsl:when>
<xsl:when test= "@type='rivedere'">

    <font color="fuchsia">
        [ <xsl:apply-templates/> ]
    </font>

</xsl:when>

```

Il foglio di stile, grazie ad una sintassi abbastanza immediata e al linguaggio XPATH, scorre l'albero della struttura implementata nella codifica XML del file in esame e applica delle trasformazioni nello stile, utilizzando, quando necessario, tag propri di HTML o persino includendo dei fogli di stile CSS. In questo caso si è utilizzato solamente XSL e come output è stato scelto il formato HTML, in modo da poter organizzare in seguito una visualizzazione a blocchi confrontabili (*frames*).

Ecco un esempio di applicazione del sopraccitato foglio di stile a questo frammento di codice XML:

Gli antichi fregi senza luci, le lunghe file dei negozi spenti, le grandi pareti dei palazzi dalle finestre chiuse e dai portoni sbarrati acquistano <seg type="ripetizione" rend="tratto cerchiato a matita verde">di colpo</seg> l'immagine di una persona amica <seg type="ripetizione" rend="tratto cerchiato a matita verde">di colpo<add place="supralinear">ma</add></seg> stravolta dalla pazzia, pronta a saltarti addosso e a sbranarti senza pietà.</p>

Quello che segue è l'esempio di applicazione del frammento di codice mostrato in precedenza. Si noti che viene individuato l'attributo "ripetizione" del tag <seg> e a questo applicata una trasformazione XSL in colore verde e carattere 4 punti; vengono individuate (e numerate) le sequenze narrative; viene individuato l'attributo "rivedere", sempre del tag <seg>. Il risultato è mostrato nella Figura 4.

Figura 4.

La donna serpente 1 Vincenzo Cerami Codifica XML a cura di Lorenzo Geri, Daniele Silvi Università degli studi Roma 3 Roma 2006 Raccolta racconti Cerami

La donna serpente

Vincenzo Cerami Indiscreto italiano 26-09-2005 Lorenzo Geri, Daniele Silvi Codifica livello 1

di VINCENZO CERAMI

I brutti quartieri di una grande città fanno quasi sempre paura. Ma a questa paura si finisce per abituarci. Quando mi trovo a passare tra quei palazzi **brutti** e violenti, ormai senza più rendermene conto, abbasso gli occhi e tiro via. Al contrario, i bellissimi quartieri del centro storico mi sono familiari, mi rassicurano a tutte le ore del giorno e in tutti i mesi dell'anno. Tranne che d'agosto, specialmente di notte.

sequenza_1

individuazione sequenza narrativa

[-] Nelle notti d'agosto il centro storico diventa addirittura terrorizzante. Gli antichi fregi senza luci, le lunghe file dei negozi spenti, le grandi pareti dei palazzi dalle finestre chiuse e dai portoni sbarrati acquistano **di colpo** l'immagine di una persona amica **di colpo** [**ma**] stravolta dalla pazzia, pronta a saltarti addosso e a sbranarti senza pietà.

Domenica sera, dopo aver passato tutta la giornata in casa a finire un lavoro che mi aveva costretto a rimanere [**nella**] **in questa** città abbandonata [**alle**] **nelle** grinfie del caldo, uscii a piedi, a caccia di un ristorante. Volevo fare due passi per sgranchirmi la schiena e anche perché avevo voglia di vedere qualche faccia vivente. Sono infatti un appassionato di facce: mi piace guardarle e immaginare [**da quali ambienti vengono fuori, quali pensieri passano per quelle teste, mi piace figurarmi la casa e la vita delle persone**] sconosciute. Mi resi conto che era domenica perché nel raggio di un chilometro almeno da casa non trovai aperta nemmeno una pizzeria. Mi perdevo per le strade attratto dalle luci lontane e nel camminare, quasi automaticamente, tornavo con i pensieri al lavoro che avevo lasciato. Così, dopo un po' di tempo, risvegliandomi dal torpore, mi resi conto di essermi perso. Mi muovevo nelle vicinanze del centro storico tra palazzi primi Novecento, quasi tutti uguali: un quartiere con pochi negozi e dalla fioca illuminazione. Le luci lontane si rivelavano puntualmente **per** scritte pubblicitarie rimaste inutilmente accese. Intanto già accusavo un po' di stanchezza alle gambe. Ma quel che più sentivo e mi cresceva dentro era una strana tensione, forse una paura. Tutti quei palazzi spenti, quegli appartamenti deserti accendevano in me impercettibili e irrazionali istinti vandalici, quasi delinquenziali. Mi passò per la testa un'improvvisa fantasia. Immaginai di imbattermi, proprio all'angolo della strada, in una **bellissima** ragazza, come me sperduta nella città notturna e afosa. Avrei resistito a non fermarla per un braccio? Mi tornò in mente un raccontino che lessi tanti anni fa. Vi si narrava proprio di un uomo sperduto nella notte che incontra una ragazza **bellissima**, avvolta in un abito bianco, di veli, un corpo magnetico, due occhi di smeraldo. L'uomo la fissa incredulo e la ragazza lo invita a seguirlo per quelle stradine medievali: qualche ramarro ai muri, qualche topo che fugge. La ragazza entra in una vecchia casa e comincia a salire le scale girandosi ogni tanto con un sorriso. Apre la porta di un appartamento sontuoso e cadente, tappezzato di arazzi e damaschi. Lui, a piccoli passi, le va dietro. La fanciulla gira l'interruttore della luce e nel soffitto di una grandissima camera da letto s'accende un lampadario di cristallo. Il letto è grande, a baldacchino. Lei sposta le lenzuola e si lascia scivolare **su** [**fra**] quelle sete. Guarda l'uomo e lo invita

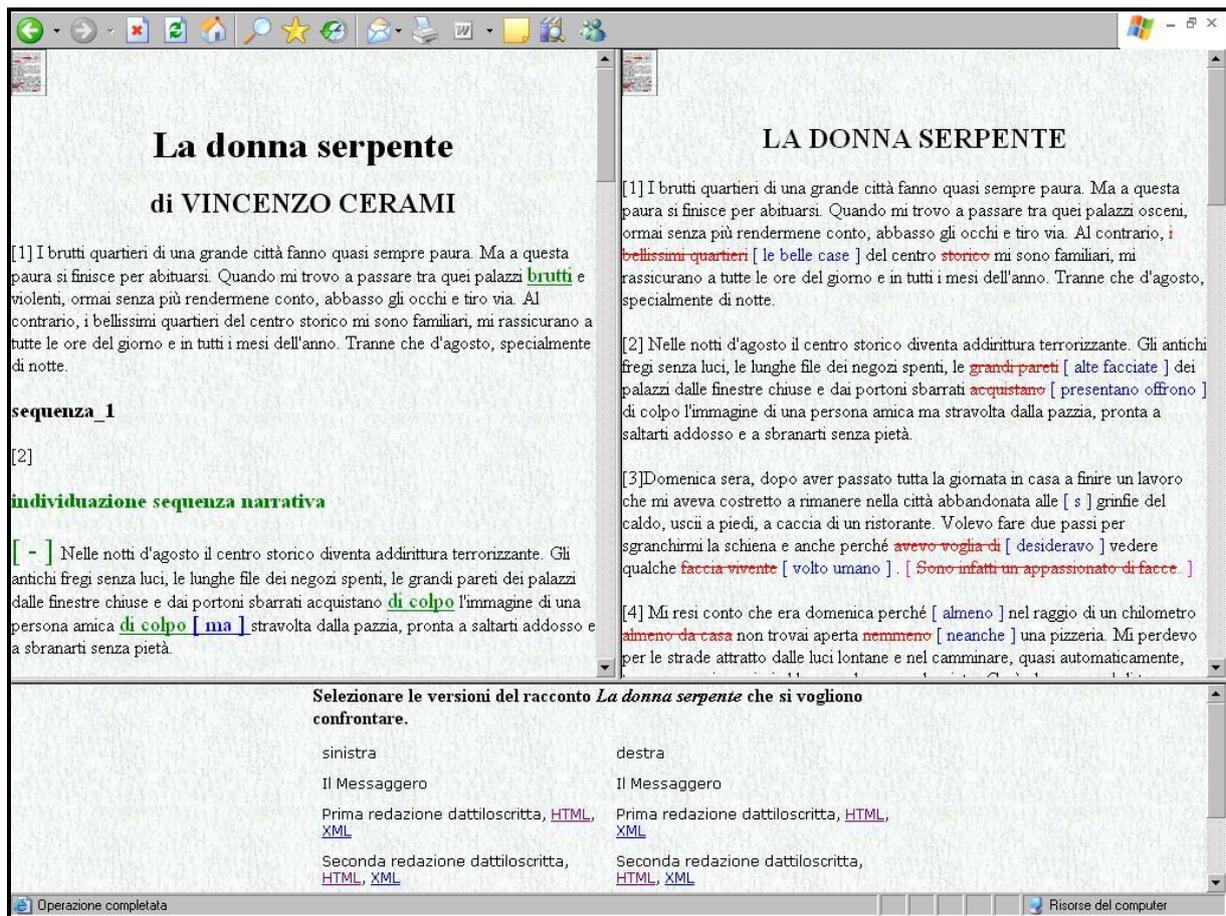
Operazione completata Risorse del computer

Notiamo come entrambe le operazioni di codifica – XML e XSL – siano di fatto atti ermeneutici: il concetto di “codifica come atto interpretativo” (Ciotti 2005) è infatti valido anche per XSL, in quanto la scelta di “trasformare” un determinato elemento del testo graficamente in una maniera piuttosto che in un’altra genera nel lettore un impatto diverso sul testo stesso e sulla sua genesi (trattandosi, in questo caso, dell’analisi di una struttura di varianti), oltre a far emergere il diverso punto di vista del codificatore. Di fatto si sceglie di modellizzare a sua volta un modello (la codifica XML) e quindi operare delle scelte sulle sequenze o sui fenomeni (precedentemente rilevati) che si ritengono rilevanti o maggiormente significativi rispetto ad altri. Dunque da un lato influisce sulla dimensione di fruizione del testo (l’interfaccia), dall’altro opera, per così dire, una ‘rappresentazione della rappresentazione’. È chiaro che di fronte a una stratificazione testuale maggiore aumenterà la profondità della codifica e dunque la difficoltà a realizzare fogli di stile in grado di visualizzare i fenomeni scrittori più complessi.

Nel nostro esempio (Figura 4) uno dei vantaggi è che la visualizzazione rende immediatamente visibili alcuni aspetti che con la semplice codifica possono passare inosservati, come la vicinanza, la frequenza o la sistematicità di certe ripetizioni.

Il tutto è stato poi assemblato in una struttura a *frames* di HTML, ottenendo il risultato che si osserva nella Figura 5.

Figura 5.

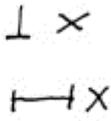


5. Codifica della correzione di bozze

Un caso particolare della codifica del dossier genetico di Vincenzo Cerami, al quale accenneremo brevemente, è costituito dalla correzione delle bozze, ultima fase del processo di revisione dell'autore in vista della pubblicazione. Se è vero che si può considerare la correzione di bozze alla stregua di ogni altra correzione, è altresì vero che si tratta di una modalità correttoria diversa dalla correzione dei dattiloscritti, per le seguenti ragioni:

- quando l'autore corregge le bozze non dialoga con se stesso, bensì con un tecnico che tradurrà *alla lettera* le sue indicazioni sulla pagina stampata
- la correzione delle bozze è l'ultima occasione, per l'autore, di rivedere il suo testo: correggendo le bozze l'autore si rivolge già ai futuri lettori.
- la correzione delle bozze avviene utilizzando simbologie standard (che in inglese è chiamata proprio *markup*).

Abbiamo ritenuto quindi di codificare le bozze in un modo diverso dalle correzioni autografe sui manoscritti. Anche in questo caso abbiamo deciso di utilizzare la TEI P4, gli attributi con `` e `<add>` insieme all'attributo *type*. Il valore da associare a *type* è una sigla che si riferisce alle azioni previste da una tabella standard per i correttori di bozze (UNI 5041, *Correzione delle bozze di stampa* 1969). Ad esempio nella tabella UNI 5041 l'azione "sostituire lettere o parole" è indicata nel seguente modo:

Eliminare lettere o parole	
----------------------------	---

Nella nostra codifica tale azione è indicata nel seguente modo:

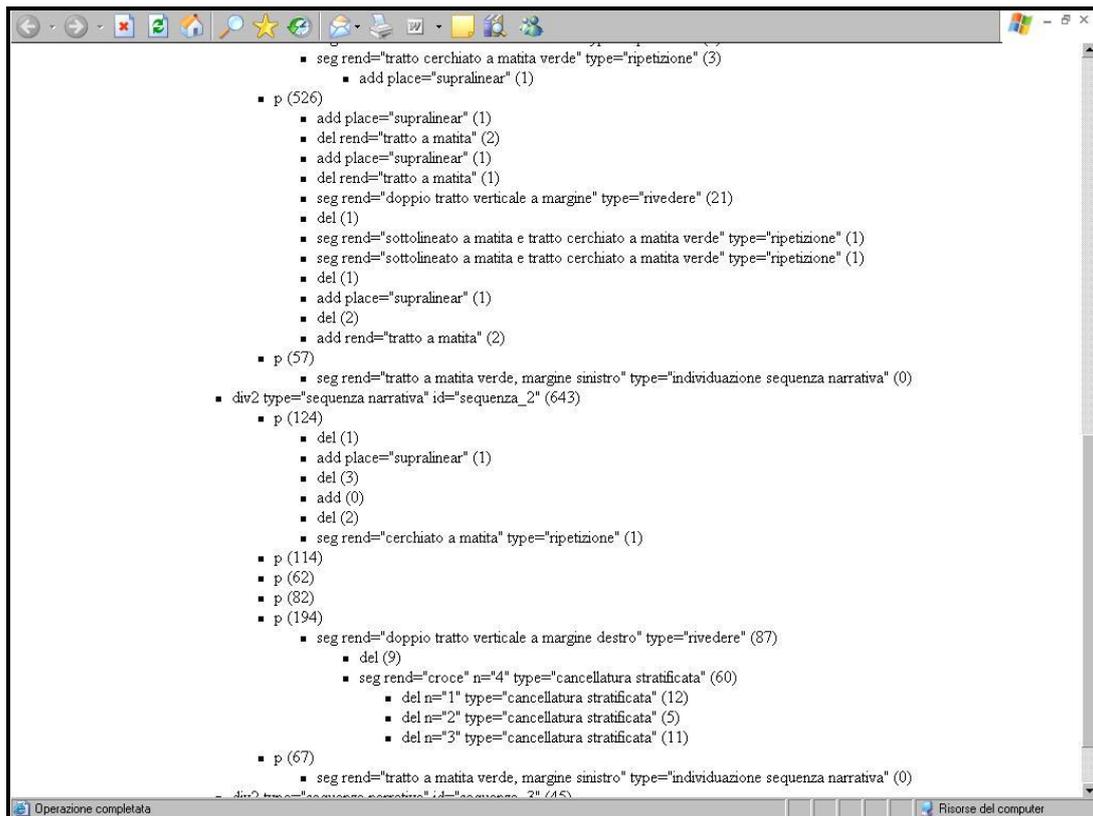
sostituire lettera	<code><seg type="cerami_bozze_sla"></code>
sostituire lettere	<code><seg type="cerami_bozze_sle"></code>

6. Un esperimento di analisi automatica con TAPoR

Ricorrere a una codifica che sia non solo descrittiva, ma anche semantica, permette di affinare l'analisi stilometrica e lessicale delle diverse versioni dei testi in esame. Per sperimentare le nuove potenzialità di analisi generate dalla codifica abbiamo pensato di sottoporre un campione dei nostri testi all'analisi automatica attraverso il software online TAPoR (<http://strange.mcmaster.ca/~taporware/>), acronimo di *Text Analysis Portal for Research*. Questo strumento per l'analisi testuale, ideato e realizzato dalla collaborazione di sei università canadesi (McMaster, Università di Victoria, Università di Alberta, Università di Toronto, Università di Montreal e Università di New Brunswick) prevede la creazione di un'unica infrastruttura a livello nazionale, formata dall'unione delle ricerche di tutte le università citate. Il portale rappresenta una via d'accesso a strumenti per sofisticate analisi e recupero di informazioni, corredato da testi campione per la sperimentazione, vale a dire testi di storia, diritto, letteratura, ecc. Oltre a fornire un'ulteriore strumento per l'analisi testuale, TAPoR è un'opportunità per la comunità degli umanisti per sviluppare modelli, strumenti e metodi che potrebbero influenzare l'evoluzione della rete, come per esempio il progetto del Semantic Web del W3C (<http://www.w3.org/DesignIssues/Semantic.html>).

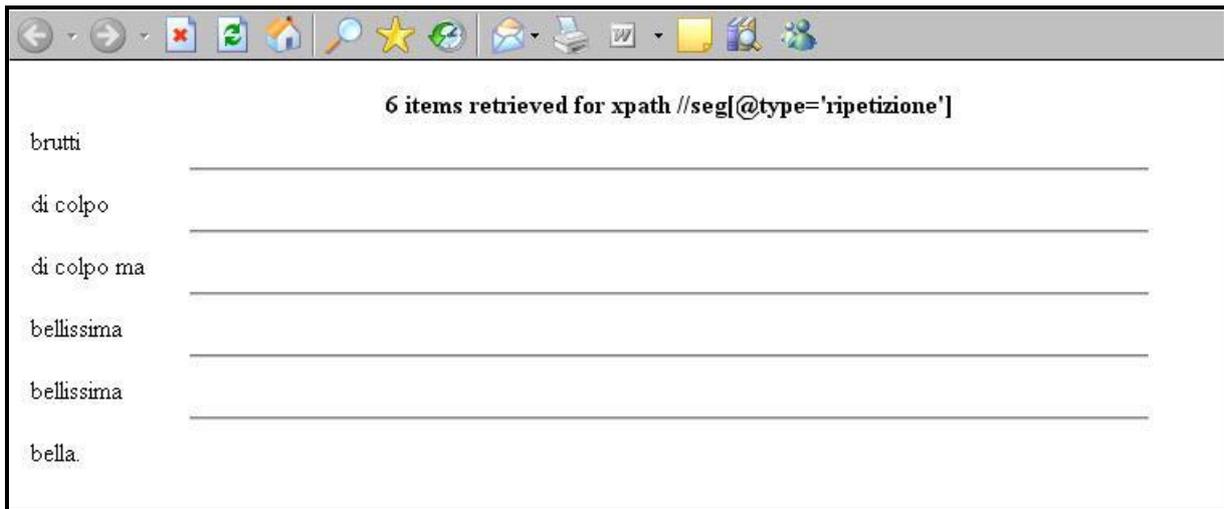
TAPoR analizza tre tipologie di testi: plain text (formato .txt), HTML, XML. In questa sede mostreremo solo alcune analisi puramente esemplificative, richiedendo un simile lavoro una trattazione più specifica. Quello che è importante notare è la possibilità di fare ricerche testuali servendosi della struttura fornita dai tag di XML. La Figura 6 (risultato di un'interrogazione con TAPoR), mostra la struttura dei tag utilizzati nella codifica di una delle versioni del racconto di Cerami¹⁰, oggetto di questo contributo.

Figura 6.



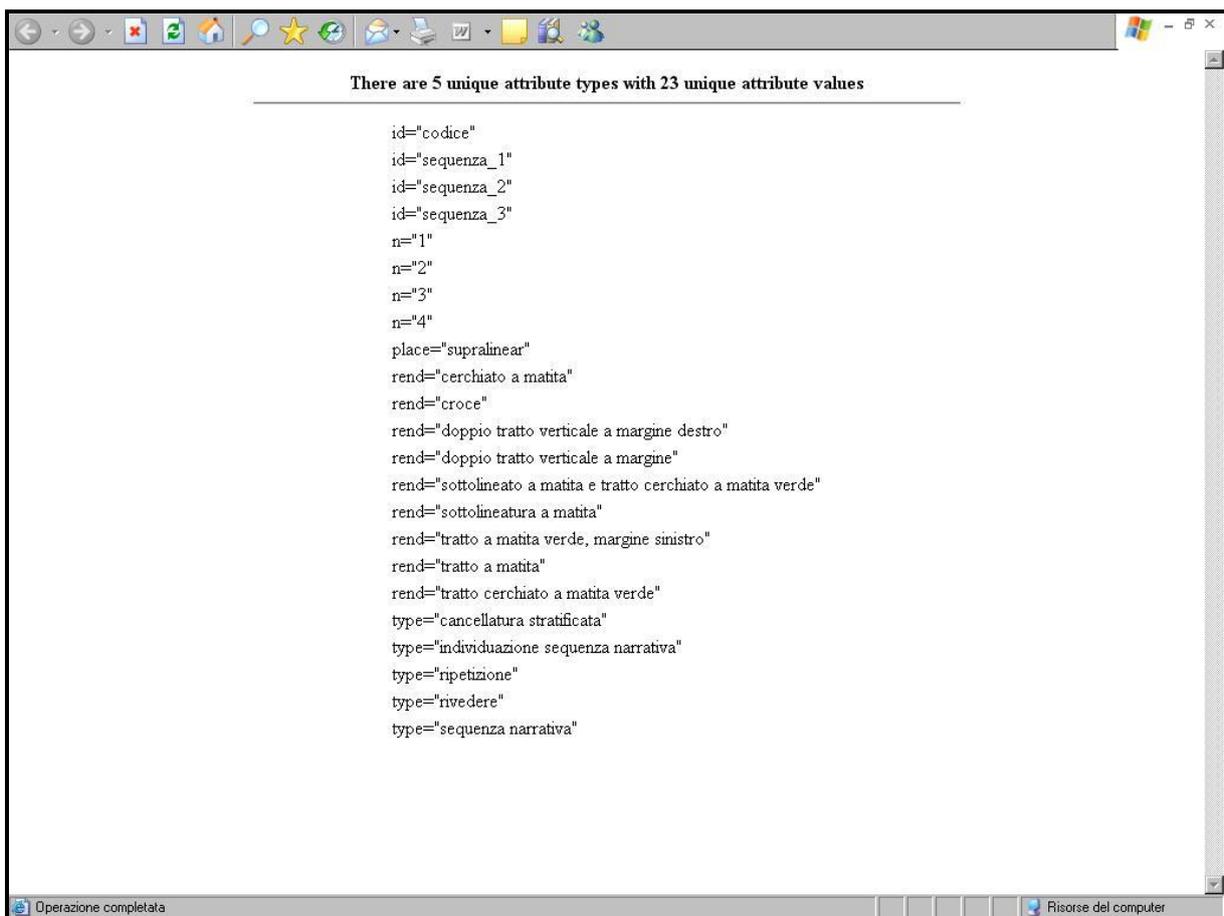
Possiamo poi concentrare la nostra attenzione sulle ripetizioni e chiedere quindi al programma di contare tutti i brani di testo contrassegnati come "ripetizioni". Si noti che in questo caso TAPoR utilizza proprio XPATH e permette quindi di indagare anche il valore degli attributi (Figura 7).

Figura 7.



Con lo stesso sistema possiamo anche vedere quanti e quali attributi sono stati utilizzati in una determinata codifica (Figura 8)

Figura 8.



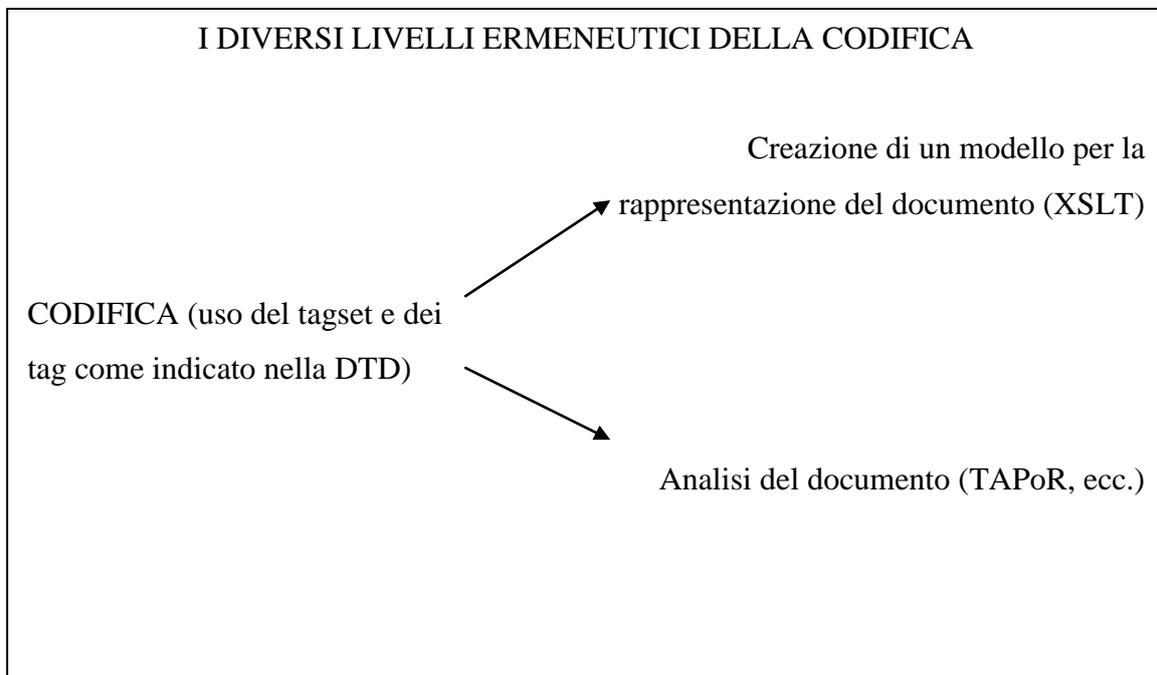
Come si può vedere l'analisi condotta in maniera capillare sia sui tag che sugli attributi ci fornisce l'opportunità di:

(a) evidenziare i cambiamenti (di qualsivoglia natura) all'interno dei singoli episodi (o sequenze narrative) in una stessa redazione;

(b) evidenziare come gli episodi cambino da una redazione all'altra e in che diverso ordine si dispongano in ogni singola redazione.

Riassumendo le considerazioni fatte in questo paragrafo e in quello precedente possiamo dire che l'atto di codifica XML apre due strade interpretative diverse e complementari tra loro, quella della trasformazione XSL e quella dell'analisi statistica. Entrambe le operazioni sono strati della 'macchina ermeneutica' messa in moto dalla codifica (vedi Figura 8).

Figura 8.



7. Opere citate

- Albonico, S.** (2005). Sull'utilizzo della codifica TEI in filologia. In Burnard, L. and Sperberg-McQueen, M., *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari*. Milano: Edizioni Sylvestre Bonnard, pp. 239-256.
- Buzzetti, D.** (2002). Il sogno e la precisione. In Mordenti, R. (ed.), *Giuseppe Gigliozzi: la fondazione dell'informatica applicata al testo letterario*. Numero monografico di *Testo & Senso*, 4/5: 24-28.
- Calquist, J.** (2004). Medieval Manuscripts, Hypertext and Reading. Vision of Digital Editions, *Literary and Linguistic Computing*, 19. 105-118.
- Ciotti, F.** (2005). La codifica del testo, XML e la Text Encoding Initiative. In Burnard, L. and Sperberg-McQueen, M., *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari*. Milano: Edizioni Sylvestre Bonnard, pp. 9-42.
- Fiormonte, D.** (1995). Varianti elettroniche, *Italiano & Oltre*, 2: 87-94.
- Fiormonte, D.** (1996). Il computer e la scrittura: forme e limiti di un influsso. In Ricciardi, M. (ed.), *Lingua Letteratura Computer*. Torino: Bollati Boringhieri, 1996, pp. 88-123.
- Fiormonte, D.** (2003). *Scrittura e filologia nell'era digitale*. Torino: Bollati Boringhieri
- Fiormonte, D. and Martiradonna, V.**, (2006). *La codifica digitale come atto ermeneutico e semiotico. Il caso di Valerio Magrelli*. In Cazalé, C. (ed.). Atti del convegno internazionale *Mémoire des textes – Textes de mémoire*, Université Paris X – Nanterre, 21-22 ottobre 2005 (in corso di pubblicazione).
- Hillesund, T.** (2005). Digital Text Cycles: From Medieval Manuscripts to Modern Markup, *Journal of Digital Information*, Volume 6 Issue 1, Article No. 309, 2005-03-09. <<http://jodi.tamu.edu/Articles/v06/i01/Hillesund/>>. Ultimo accesso: 20 novembre 2006.
- McGann, J.** (2005). From Text to Work. Digital Tools and the Emergence of the Social Text, *Variants*, 4: 225-240
- Lecolinet, E., Robert, F. and Role F.** (2002). Text-image coupling for editing literary sources, *Computers and the Humanities*, 36: 49-73.
- Neyt, V.** (2006). Fretful Tags Amid the Verbiage: Issues in the Representation of Modern Manuscript Material, *Literary and Linguistic Computing*, 21: 99-111.
- Orlandi, T.** (1999). Ripartiamo dai diasistemi. in *I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici*. Atti del convegno Internazionale dell'Accademia Nazionale dei Lincei in collaborazione con l'Associazione

Internazionale per gli Studi di Lingua e Letteratura Italiana, Roma, 27-29 maggio 1998. Roma: Accademia Nazionale dei Lincei, pp. 87-101.

Schmidt, D., (2006b). Graphical Editor for Manuscripts, *Literary and Linguistic Computing*, 21: 341-351.

Schmidt, D., (2006b). What future for digital literature? (in corso di pubblicazione).

Shillingsburg, P. L. (2006). *From Gutenberg to Google. Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press.

Vanhoutte, E. (2003). Display or Argument: Markup Visualisation for Electronic Scholarly Editions. In *Standards und Methoden der Volltextdigitalisierung*. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8/9 Oktober 2001. Hrsg. von T. Burch, J. Fournier, K. Gärtner und A. Rapp. Stuttgart: Franz Steiner Verlag, pp. 71-96.

Van Hulle, D. (2004). Pro Paralipomena: multiple versions and the importance of the documentary context. Unpublished paper given at the first *European Society for Textual Scholarship* conference, Alicante, 25-27 November 2004.

Vetter, L. and McDonald, J. (2003). Witnessing Dickinson's Witnesses, *Literary and Linguistic Computing*, 18: 151-165.

Wardrip-Fruin, N. (2005). Playable Media and Textual Instruments. <<http://www.brown.edu/Research/dichtung-digital/2005/1/Wardrip-Fruin/>>. Ultimo accesso: 20 novembre 2006.

¹ Questo contributo nasce dalla ricerca "Codifica e analisi informatica di testi letterari" finanziata dal Dipartimento di Italianistica dell'Università Roma Tre e diretta da D. Fiormonte. All'interno di una concezione e di una ricerca unitarie, i paragrafi 1-2 e la revisione generale del testo si devono a D. Fiormonte, il paragrafo 3 a L. Geri e i paragrafi 4-6 a D. Silvi.

² Tra i progetti di banche dati testuali, ne segnaliamo tre che utilizzano XML per codificare tipologie complesse di varianti in manoscritti di autori moderni e contemporanei: *Dickinson Electronic Archive* (<http://www.emilydickinson.org/>), *Henrik Ibsen's Writings* (<http://www.ibsen.uio.no/his/hjemmeside/english.html>), *The Wittgenstein Archives at the University of Bergen (WAB)* (<http://wab.aksis.uib.no/index.page>) Sulla complessità dell'edizione elettronica di manoscritti contemporanei si vedano Van Hulle 2004 (su James Joyce), Vetter / McDonald 2003 (su Emily Dickinson) e Fiormonte / Martiradonna 2006 (su Valerio Magrelli). Tra i contributi dedicati a fare il punto sulle edizioni digitali si segnalano Neyt 2006 per i manoscritti moderni e Calquist 2004 per l'età medievale. Sui limiti di XML-TEI per la rappresentazione della genesi testuale si vedano Vanhoutte 2003 e Schmidt 2006a.

³ http://www.tei-c.org/Activities/SIG/Manuscript/mssigr01.xml.ID=body.1_div.2.

⁴ Per una prospettiva sulle varianti dei testi antichi e una riflessione generale sul problema della codifica dei testi stratificati cfr. Albonico 2005.

⁵ <http://www.selc.ed.ac.uk/italian/digitalvariants>. Il progetto è analizzato e descritto in dettaglio in Fiormonte, 2003: 209-237.

⁶ Per un esperimento di codifica XML-TEI dei vari passaggi di scrittura di alcuni componimenti di Magrelli cfr. Fiormonte / Martiradonna 2006.

⁷ Il lavoro è stato realizzato nell'anno accademico 2001-02 all'interno di un corso di 30 ore di Informatica umanistica svolto da D. Fiormonte presso l'Università di Roma "La Sapienza".

⁸ Una descrizione completa dei materiali dell'archivio Cerami è in Fiormonte 1996.

⁹ La matita è la stessa quindi si può ipotizzare che si tratta di un'azione svolta poco dopo aver segnato l'episodio.

¹⁰ Si tratta della redazione dattiloscritta successiva a quella pubblicata sul *Messaggero*. In sostanza il primo testimone dattiloscritto.