

LL

12th International Colloquium on Latin Linguistics
Tavola rotonda:
*From manuscript to digital text. Problems of
interpretation and markup*
12 GIUGNO 2003

**XML/TEI per la trascrizione
delle fonti primarie e la codifica
dell'apparato**

14/06/2006 Francesca Tomasi

Nuove tecnologie e scienze umane

- **congerie di applicazioni** che si possono impiegare per l'implementazione di prodotti digitali;
- **molteplicità degli aspetti del testo** che si intendono indagare in fase di trasferimento in MRF.

Necessita di definire gli **OBIETTIVI** della resa computazionale

14/06/2006 Francesca Tomasi

Linguaggi di codifica

Scelta dell'XML (*Extensible Markup Language*):

- portabilità, durata nel tempo, autonomia dal software
- libertà interpretativa delle istanze testuali e documentarie.

14/06/2006 Francesca Tomasi

La modellizzazione

- individuazione di un **modello** del testo conforme agli obiettivi dell'indagine.
- scelta di un **linguaggio formale** (come XML) con cui implementare tale modello.
- modello derivato dalle **esigenze analitiche** operate dall'*interprete*
- codifica elaborata sulla base del modello: **operazione di interpretazione**

14/06/2006 Francesca Tomasi

Trascrizione delle fonti primarie e XML

- Fase di trascrizione delle fonti primarie: intervento interpretativo sulle istanze strutturali, semantiche e materiali trasmesse da una fonte testuale.
- XML offre ampie possibilità: elaborazione di documenti *well-formed* o sviluppo DTD (*Document Type Definition*) *ad hoc*

14/06/2006

Francesca Tomasi

XML e TEI (*Text Encoding Initiative*)

TEI (cfr. TEI P4, 2002; sito ufficiale <http://www.tei-c.org/>).

- schema divenuto oggetto di studi decennali
- completezza dei marcatori previsti

Esigenza di un sistema standard di codifica delle fonti primarie anche per la mancanza di criteri uniformi di rappresentazione dei testi manoscritti

14/06/2006

Francesca Tomasi

Struttura della TEI

Divisione dello schema di codifica in tre grandi blocchi: *core tag set*, *base tag set*, *additional tag set*, fra cui:

- TRASCRIZIONE DELLE FONTI PRIMARIE:** *TEI.transcr* ([teitran2.dtd](#) e [teitran2.ent](#))
- CODIFICA DELL'APPARATO CRITICO:** *TEI.textcrit* ([teitc2.dtd](#) e [teitc2.ent](#))

14/06/2006

Francesca Tomasi

TRASCRIZIONE DELLE FONTI PRIMARIE

Si possono individuare varie sezioni di interesse. Le più interessanti riguardano:

- 1. la codifica delle alterazioni del testo** (abbreviazioni, cancellazioni, riscritture);
- 2. la codifica di fenomeni "non linguistici"** (identificazione di mani diverse, punti illeggibili, danneggiati o deteriorati, inusuali spaziature nel testo)

14/06/2006

Francesca Tomasi

Abbreviazioni e scioglimento di abbreviazioni

<abbr> codifica un'abbreviazione di qualsiasi tipo
<expan> codifica lo scioglimento di un'abbreviazione

Possono essere usati uno come attributo dell'altro, a seconda si intenda trascrivere il testo abbreviato o lo scioglimento del testo. L'attributo **type** può essere poi usato per specificare il tipo di abbreviazione (nota tironiana, troncamento, contrazione, ecc.). Le abbreviazioni andrebbero indicate sottoforma di riferimento di entità, le cui modalità di scioglimento o visualizzazione viene specificato nella dichiarazione di entità della DTD (p con l'asta inferiore tagliata sta per "per", la linea soprascritta sta per la lettera "n", ecc.).

Es.
 <abbr expan="non">&n-trattosovrascritto;</abbr>
 <expan abbr="&p-astainferiore;">per</abbr>

14/06/2006

Francesca Tomasi

Abbreviazioni tramite riferimenti di entità

Questo stesso sistema di impiego dei riferimenti di entità può essere utile anche per codificare alcuni caratteri che identificano il sistema scrittoria di un autore, vale a dire alcuni segni di scrittura particolari. Questo significa che un riferimento di entità può essere associato ad un'immagine e quindi sostituito, in fase di *layout*, con l'immagine della lettera. Per esempio possiamo dire che *&pro;* venga sostituito, in fase di visualizzazione con la corrispettiva immagine digitale

14/06/2006

Francesca Tomasi

Esempi di abbreviazioni

Immagine delle più comuni abbreviazioni dei manoscritti latini medievali.
 Tratto da G. Cencetti, *Paleografia latina*, II ed. Jouvence, Roma 1997, p. 163. Cfr. Anche A. Cappelli, *Dizionario di abbreviature latine ed italiane*, ristampa, Hoepli, Milano 1998.



14/06/2006

Francesca Tomasi

Correzioni e congetture

<sic> codifica una porzione di testo che viene riprodotta anche se in apparenza scorretta.
<corr> codifica la forma corretta di un passaggio apparentemente errato.

Anche in questo caso possono essere usati uno come attributo dell'altro, a seconda della versione del testo si intenda riprodurre. In realtà più che di congetture, come vorrebbe la TEI, si può parlare di semplice errore del copista che viene emendato in fase di trascrizione.

Perché io voglio che quel <sic corr="diavolo">diavlo</sic>
 Perché io voglio che quel <corr sic="diavlo">diavolo</corr>

Tratto da Giovanni Sabadino degli Arienti, *Novelle Porrettane*, cod. Palatino 503, Biblioteca Nazionale Centrale di Firenze

14/06/2006

Francesca Tomasi

Aggiunte e cancellazioni

<add> codifica lettere, parole o frasi inserite nel testo dall'autore, dal copista, da un correttore, da un glossatore o annotatore. L'attributo **place** può essere utilizzato per specificare la posizione dell'aggiunta (inline, supralinear, infralinear, left, right, top, bottom)

**** codifica lettere, parole o frasi cancellate nel testo, o segnate come da cancellare, o comunque indicate come superflue dall'autore, dal copista, da un correttore, da un glossatore o annotatore. L'attributo **rend** può essere utilizzato per indicare il tipo di cancellazione (erasure, bracketed, overstrike, subpunction)

[Esempio di aggiunta marginale e di cancellazione \(testo espunto e depennato\)](#)

14/06/2006

Francesca Tomasi

Sostituzioni

Spesso cancellazioni o errori vengono rimpiazzati dal copista o da altra mano con altre stringhe di testo; questo fenomeno può essere codificato con:

<sic> e **<corr>**, da utilizzare l'uno gerarchicamente dipendente dall'altro per codificare una sequenza di sostituzioni;

**** e **<add>**, usati in sequenza per evidenziare che una porzione di testo è stata prima cancellata e poi è stato aggiunto altro testo.

[Esempio di cancellazione e aggiunta](#)

14/06/2006

Francesca Tomasi

Ripristino di porzioni in precedenza cancellate e altri segni di evidenziazione

Un autore può decidere di evidenziare una parola o una frase in qualche modo e poi riflettere di cancellare l'evidenziazione. Ad esempio una porzione di testo può essere evidenziata come da cancellare e poi la cancellazione essere eliminata, e quindi provvedere al ripristino del testo prima cancellato.

L'elemento **<restore>** indica il ripristino di un testo ad uno stadio precedente la cancellazione o altro segno di evidenziazione o istruzione di modifica.

14/06/2006

Francesca Tomasi

Testo omesso nella trascrizione

<gap> indica un punto del testo una porzione non è stata inserita nella trascrizione elettronica (per esempio in quanto illeggibile, oppure perché mancante). L'attributo **desc** può essere utilizzato per spiegare la ragione dell'omissione, **reason** fornisce la ragione dell'omissione mentre l'attributo **extent** consente di specificare la dimensione della omissione (esprimibile in numero di lettere o in centimetri).

<supplied> indica la presenza di una stringa di testo come possibile integrazione della lacuna (completamente congetturale). L'attributo **reason** consente di specificare le ragioni dell'impossibilità di leggere

[Esempio di lacuna illeggibile](#)

14/06/2006

Francesca Tomasi

Altri elementi rilevanti

<hi> evidenzia una parola o una porzione graficamente distinta dal testo circostante (per esempio un'iniziale rubricata).
L'attributo per specificare le caratteristiche dell'evidenziazione è **rend**

[Clicca per l'esempio](#)

14/06/2006

Francesca Tomasi

Attributi universali

Esiste poi l'attributo **resp**, o l'attributo **hand**, che consente di assegnare la paternità a certe operazioni, vale a dire di definire chi è il responsabile di certe azioni sul testo, delle interpretazioni o del *markup*: la mano per un'aggiunta o una cancellazione, lo scioglimento di un'abbreviazione, la correzione di un errore apparente, la trascrizione di porzioni non chiare, la decisione di non trascrivere alcune porzioni di testo.

14/06/2006

Francesca Tomasi

Fenomeni non linguistici

Punto in cui cambia la mano dello scriba, con quali caratteristiche varia la scrittura (diverso inchiostro o diverso strumento di scrittura), punti in cui il testo è danneggiato o illeggibile, spazi inusuali o linee nella fonte.

14/06/2006

Francesca Tomasi

Identificazione e descrizione delle mani diverse nei documenti

Elemento **<hand>**

Usato nell'intestazione **<teiHeader>** (nella sezione del **<profileDesc>**) per distinguere ogni scriba e ogni diverso stile di scrittura. Le indicazioni delle varie mani (cioè ogni elemento **<hand>**) vanno inserite all'interno del marcatore **<handList>** (cioè lista delle varie mani)

14/06/2006

Francesca Tomasi

Attributi utilizzabili per identificare le varie mani

id - come identificativo univoco usato poi nel documento per riferirsi ad uno scriba o ad uno stile di scrittura;
 type - per identificare il tipo di scrittura
 scribe - per assegnare il nome o l'identificativo di ogni scriba;
 style - per indicare lo stile di scrittura;
 ink - per descrivere il colore dell'inchiostro;
 character - per descrivere altre caratteristiche della mano;
 resp - per indicare il nome dell'editore o del trascrittore responsabile per l'identificazione della mano.

[Clicca per esempio](#)

14/06/2006

Francesca Tomasi

Codifica del cambio di mano nel testo

Ogni punto del testo in cui la mano cambia viene codificato con l'elemento vuoto **<handShift>** accompagnato dagli attributi **old** e **new** per indicare il codice identificativo della precedente mano e il codice della successiva (codici che corrispondono agli **id** che accompagnano le varie **hand** specificate nell'**handList**)

[Esempio di cambio di ductus](#)

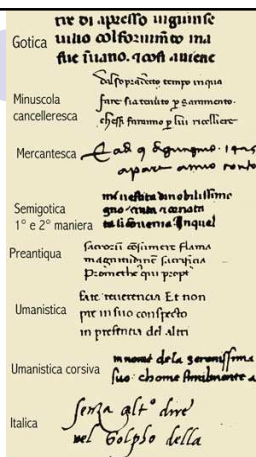
14/06/2006

Francesca Tomasi

Alcuni esempi di scritture.

L'evoluzione dei sistemi scrittori.

Tratto da A. Petrucci,
Breve storia della scrittura latina,
 nuova ed. Bagatto,
 Roma 1992, p. 206.



14/06/2006

Francesca Tomasi

Danni, porzioni illeggibili, testi aggiunti

Assieme a **<gap>** e **<supplied>** (che si utilizzano quando il testo è completamente illeggibile) anche **<damage>** fa riferimento ad un'area di testo danneggiata che però è leggibile.

Fra gli attributi **type**, che permette di specificare la tipologia di danno (abrasione, macchia, ecc...), **agent** che consente di specificare, quando possibile, l'agente che ha causato il danno (attacco microbico, foxing) e **extent** che permette di specificare l'estensione del danno (espresso in numero di lettere, centimetri o porzioni strutturali).

Es. `<damage type='macchia' agent='foxing' extent='3 lettere'/>`

<unclear> codifica una parola, una frase o un passaggio che non viene trascritto con certezza in quanto illeggibile.

Questi elementi possono anche essere utilizzati uno all'interno dell'altro:
`<damage type='macchia' agent='foxing' extent='3 lettere'>
 no<unclear>str</unclear>um</damage>`

[Esempio di danno con testo leggibile](#)

14/06/2006

Francesca Tomasi

Spazi

Lo scriba o il copista può aver lasciato uno spazio bianco nel testo (per un'iniziale miniata o in quanto non riusciva a leggere dalla fonte o per qualche altra ragione). L'elemento **<space>** consente di rilevare la presenza di una porzione bianca e l'attributo **extent** di definirne le dimensioni (in numero di lettere o centimetri). L'elemento **<supplied>** può essere utilizzato per integrare la mancanza.

[Esempio](#)

14/06/2006

Francesca Tomasi

Resa in layout

Sarà poi giudizio del codificatore decidere come utilizzare XSLT (*Extensible Stylesheet Language Transformation*, linguaggio di visualizzazione di XML sul browser) per rendere in layout i vari fenomeni codificati, ricorrendo a strategie di visualizzazione e resa grafica.

Vediamo un [esempio](#), tratto dalle Novelle Porrettane di Giovanni Sabadino degli Arienti (cod. Palatino 503, c. 64) di applicazione del linguaggio XML (TEI) per la codifica dei segni del manoscritto.

14/06/2006

Francesca Tomasi

L'apparato critico

L'apparato è il luogo destinato a documentare le scelte compiute in fase di definizione del testo critico. L'apparato riporta cioè le varianti riprodotte nei testimoni della tradizione testuale. Tutti gli elementi visti per la trascrizione delle fonti primarie possono essere utilizzati anche nell'apparato per codificare le varianti dei codici attestanti la tradizione testuale.

Le informazioni relative possono essere inserite all'interno del testo codificato e poi visualizzate secondo strategie variabili in sede di layout

14/06/2006

Francesca Tomasi

Elementi per la codifica dell'apparato

L'elemento **<app>** contiene le varie lezioni ricondotte dai testimoni della tradizione. L'elemento **<lemm>** riporta la lezione accettata nel testo critico, **<rdg>** (reading) ogni singola lezione variante. Questo elemento supporta l'attributo **wit** (witness) che, come valore, riporta la sigla identificativa del testimone che tramanda la lezione variante ed eventualmente l'attributo **type** che consente di specificare la tipologia della variante (errore di trascrizione, correzione o intervento sostitutivo).

14/06/2006

Francesca Tomasi

Esempio

```
<app>
  <lemm wit='R'>avocarem</lemm>
  <rdg wit='M'>avocam</rdg>
  <rdg wit='F'>evocarem</rdg>
</app>
```

Es. Tratto da Poggio Bracciolini, Lettere, vol. I Lettere a Niccolò Niccoli, a cura di H. Harth, Olscki, Firenze 1983, p. 109, riga 4.

14/06/2006

Francesca Tomasi

Elenco dei testimoni

La serie delle informazioni sui vari testimoni della tradizione può essere racchiusa all'interno dell'elemento **<witList>**, che contiene la lista dei testimoni, ciascuno codificato con l'elemento **<witness>**. L'attributo **sigil** permette di assegnare una sigla identificativa per ogni testimone (con la quale viene poi richiamato nel testo).

14/06/2006

Francesca Tomasi

Esempio

```
<witList>
  <witness sigil="R">Firenze, Biblioteca
  Riccardiana, 759</witness>
  <witness sigil="M">Venezia, Biblioteca
  Marciana, Lat. XI, 66</witness>
  <witness sigil="F">Ferrara, Biblioteca
  Comunale, Cl. II, 133</witness>
  ...
</witList>
```

Poggio Bracciolini, Lettere, vol. I Lettere a Niccolò Niccoli, a cura di H. Harth, Olscki, Firenze 1983.

14/06/2006

Francesca Tomasi