

Manuale di codifica XML/TEI

La DTD TeiLite Introduzione alla codifica dei testi

Lou Burnard - C. M. Sperberg-McQueen
Documento N. TEI U5, Giugno 1995
Traduzione italiana di: Fabio Ciotti, Guendalina Demontis, Giuseppe Gigliozzi,
Massimo Guerrieri, Andrea Loreti
Revisione e cura traduzione italiana di: Fabio Ciotti
Gennaio 1998
Revisione versione XML a cura di: Francesca Tomasi



Principi generali

Questo documento fornisce un'introduzione alle indicazioni elaborate dalla *Text Encoding Initiative* (TEI), descrivendo un sottoinsieme facilmente utilizzabile dell'intero schema di codifica. Lo schema qui documentato può essere utilizzato per codificare una vasta gamma di caratteristiche testuali comunemente riscontrate, in modo da ottimizzare l'utilizzabilità delle trascrizioni elettroniche e facilitare il loro scambio fra studiosi che utilizzano diversi sistemi informatici. Esso è altresì pienamente compatibile con l'intero schema TEI, definito dal documento *TEI P3, Guidelines for Electronic Text Encoding and Interchange*, pubblicato a Chicago e Oxford nel maggio del 1994.

- La versione originale di questo testo - in lingua inglese - può essere reperita attraverso World Wide Web all'indirizzo:
 - <http://www.tei-c.org/Lite/>
- La *Document Type Definition* formale qui descritta, si trova negli stessi siti, nel file *teilight.dtd*:
 - <http://www.tei-c.org/Lite/DTD/>



Introduzione



- Le *Norme (Guidelines)* della *Text Encoding Initiative (TEI)* sono indirizzate a tutti coloro che intendono scambiare informazioni archiviate in formato elettronico. Esse sottolineano l'importanza dello scambio di informazioni testuali, ma trattano anche di altre forme di informazioni (quali immagini e suoni). Le *Norme* sono applicabili indifferentemente sia per la creazione di nuove risorse che per lo scambio di quelle già esistenti.
- Le *Norme* forniscono un mezzo per rendere esplicite certe caratteristiche di un testo in modo tale da facilitarne il trattamento mediante programmi di computer basati su diverse piattaforme. Definiamo questo processo di esplicitazione *marcatura (markup)* o *codifica (encoding)*. Qualsiasi rappresentazione di un testo su un computer usa una qualche forma di codifica; la TEI è stata creata sia per ovviare alla eccessiva varietà di schemi di codifica tra loro incompatibili che ostacolano la ricerca scientifica, sia per il crescente numero di applicazioni scientifiche che ora vengono individuate per i testi in formato elettronico.

Sponsor



La TEI è sponsorizzata dall'*Association for Computers and the Humanities*, dall'*Association for Computational Linguistics*, e dall'*Association for Literary and Linguistic Computing*. Finanziamenti sono stati in parte forniti dall'*U.S. National Endowment for the Humanities*, *Directorate General XIII of the Commission of the European Communities*, dall'*Andrew W. Mellon Foundation*, e dal *Social Science and Humanities Research Council of Canada*.

Le *Guidelines* sono state pubblicate nel maggio 1994, dopo sei anni di sviluppo che ha coinvolto parecchie centinaia di studiosi di tutto il mondo provenienti da diversi ambiti accademici.

Principi generali



Le *Norme* dovrebbero:

- essere in grado di rappresentare le caratteristiche testuali necessarie per la ricerca;
- essere semplici, chiare e concrete;
- essere di semplice utilizzazione per i ricercatori senza il ricorso a *software* specializzati;
- permettere una definizione rigorosa e un'efficiente elaborazione dei testi;
- consentire estensioni definite dall'utente;
- essere conformi agli standard esistenti o in procinto di essere adottati.

Obiettivi del sottoinsieme TEI (Lite)



- esso dovrebbe includere la maggior parte dell'insieme 'fondamentale' di marcatori della TEI, dal momento che questo contiene elementi rilevanti virtualmente per tutti i testi e per tutti i tipi di elaborazione testuale;
- esso dovrebbe essere in grado di trattare adeguatamente una varietà di testi ragionevolmente ampia, al livello di dettaglio incontrato nella pratica già esistente;
- esso dovrebbe essere utile sia per l'elaborazione di nuovi documenti sia per la codifica di quelli già esistenti;
- esso dovrebbe essere utilizzabile con un ampio spettro di applicazioni XML già esistenti;
- esso dovrebbe essere derivabile dall'intera DTD TEI, usando il meccanismo di estensione descritto nelle *Norme*;
- esso dovrebbe essere tanto conciso e semplice quanto consentito dalla conformità agli altri obiettivi.



Struttura di un testo TEI

- Tutti i testi conformi alla TEI contengono:
 - (a) una *testata TEI* (codificata come `<teiHeader>`) e
 - (b) una trascrizione del testo vero e proprio (marcata con l'elemento `<text>`).
- La testata TEI contiene informazioni analoghe a quelle contenute nel frontespizio di un testo a stampa. Essa può avere fino a quattro parti: una descrizione bibliografica del testo memorizzato su supporto digitale, una descrizione del modo nel quale è stato codificato, una descrizione non bibliografica del testo (un profilo del testo), e una storia delle revisioni.
- Un testo TEI può essere *unitario* (un'opera singola) o *composito* (una collezione di opere, come un'antologia). In entrambi i casi, il testo può avere un *avantesto* o degli *annessi* opzionali. In mezzo c'è il *corpo* del testo, che, nel caso di un testo composito, può essere costituito da *gruppi*, ciascuno dei quali contiene più gruppi o testi.



Struttura generale TEI

Un testo unitario sarà codificato usando una struttura generale come questa:

```
<TEI.2>
  <teiHeader> [testata TEI] </teiHeader>
  <text>
    <front> [avantesto] </front>
    <body> [testo] </body>
    <back> [annessi] </back>
  </text>
</TEI.2>
```

Struttura testo composito

Anche un testo composito può essere dotato opzionalmente di avantesto e annessi.
Nel mezzo si trovano uno o più gruppi di testi, ognuno eventualmente dotato di propri avantesti ed annessi.



```
<TEI.2>
  <teiHeader> [testata del testo composito] </teiHeader>
  <text>
    <front> [avantesto del testo composito] </front>
    <group>
      <text>
        <front>[avantesto del primo testo] </front>
        <body> [primo testo unitario] </body>
        <back> [annessi del primo testo] </back>
      </text>
      <text>
        <front>[ avantesto del secondo testo] </front>
        <body> [secondo testo unitario] </body>
        <back> [annessi del secondo testo] </back>
      </text>
      [altri testi o gruppi di testi]
    </group>
    <back> [annessi del testo composito] </back>
  </text>
</TEI.2>
```

Codifica del corpo del testo

Come indicato sopra, un semplice documento TEI è costituito - a livello testuale - dai seguenti elementi:

- *<front>* contiene qualsiasi elemento (testate, frontespizio, prefazioni, dediche, etc.) che si trovi prima dell'inizio del testo vero e proprio.
- *<group>* raggruppa un insieme di testi unitari o gruppi di testi.
- *<body>* contiene l'intero corpo di un singolo testo unitario, con l'esclusione di ogni elemento preliminare o di appendice.
- *<back>* contiene qualsiasi annesso o appendice che segua la parte principale di un testo.



Elementi per la segmentazione del testo (<body>)



Il corpo di un testo in prosa può essere costituito solamente da una serie di paragrafi, oppure tali paragrafi possono essere raggruppati insieme in capitoli, sezioni, sottosezioni, etc.

Nel primo caso, ogni paragrafo è marcato usando il marcatore `<p>`. Nel secondo caso, l'elemento `<body>` può essere diviso sia in una serie di elementi `<div1>`, sia in una serie di elementi `<div>`, i quali possono essere ulteriormente suddivisi

Marcatori di segmentazione



`<p>` marca i paragrafi in prosa.

`<div>` contiene una sezione dell'avantesto, del corpo o degli annessi di un testo.

`<div1>` contiene una sezione di primo livello dell'avantesto, del corpo ed degli annessi di un testo (la più ampia, se `<div0>` non è usato, la seconda in ordine di grandezza altrimenti)

Qualora siano necessarie suddivisioni strutturali più piccole di una `<div1>`, quest'ultima può essere divisa in elementi `<div2>`, una `<div2>` in più piccoli elementi `<div3>`, etc., fino al livello `<div7>`. Se sono presenti più di 7 livelli della divisione strutturale, è necessario o modificare la DTD TEI per accettare `<div8>`, etc, oppure usare l'elemento non numerato `<div>`: un elemento `<div>` può essere suddiviso in elementi più piccoli `<div>`, senza limiti alla profondità di annidamento.

Attributi degli elementi di segmentazione



type

indica il nome convenzionale per questa categoria di divisione del testo. Il suo valore sarà generalmente *LIBRO*, *CAPITOLO*, *POEMA*, etc. Altri possibili valori includono *GRUPPO* per gruppi di poemi, etc, trattati come un'unità singola, *SONETTO*, *DISCORSO* e *CANZONE*. Si noti che qualsiasi valore venga fornito per l'attributo *type* del primo `<div>`, `<div1>`, `<div2>`, etc., in un dato testo, si assume che sia valido per tutte le successive `<div>`, `<div1>`, all'interno dello stesso `<body>`. Questo implica che un valore deve essere fornito per il primo elemento di divisione di ogni tipo, o ogni qualvolta il valore cambi.

id

specifica un identificatore unico per la divisione, che può essere usato per i riferimenti incrociati o per altri collegamenti, come un commento, come più avanti discusso nella sezione *Riferimenti incrociati e collegamenti*. Spesso è utile dare un attributo *id* a ogni unità strutturale importante di un testo, e far derivare i valori in un modo sistematico, ad esempio aggiungendo un numero di sezione a un breve codice per il titolo dell'opera in questione, come negli esempi che seguono.

n

specifica un breve nome mnemonico o un numero per la divisione, che può essere usato per identificarla, preferendolo a *id*. Se già esiste una forma convenzionale di riferimento o un'abbreviazione per le parti di un lavoro (come il libro/capitolo/verso delle citazioni bibliche), l'attributo *n* è il luogo deputato a registrarlo.

Caratteristiche degli attributi



- Gli attributi *id* e *n*, peraltro, sono tanto utili da essere ammessi in tutti gli elementi, in ogni DTD conforme alla TEI: sono *attributi globali*. Gli altri attributi globali definiti nello schema *TEI Lite* sono discussi nella sezione *Attributi di collegamento*.
- Il valore di ogni attributo *id* deve essere unico all'interno di un documento. Un semplice modo per assicurarsi che questo avvenga è fare in modo che esso rifletta la struttura gerarchica del documento.
- Per esempio, *La ricchezza delle nazioni* di Smith, nella prima edizione, consiste di cinque libri, ognuno dei quali è diviso in capitoli, alcuni dei quali sono ulteriormente suddivisi in parti.

Esempio di codifica (elemento <div> e attributi)



```
<div1 id='RN1' n='I' type='libro'>
  <div2 id='RN101' n='I.1' type='capitolo'> ... </div2>
  <div2 id='RN102' n='I.2' type='capitolo'> ... </div2>
  ...
  <div2 id='RN110' n='I.10' type='capitolo'>
    <div3 id='RN1101' n='I.10.1' type='sezione'>... </div3>
    <div3 id='RN1102' n='I.10.2' type='sezione'>... </div3>
  </div2>
  ...
</div1>
<div1 id='SN2' n='II' type='libro'>
  ...
</div1>
...
```

Titoli e chiusure



Ogni <div>, <div1>, <div2>, etc., può avere un titolo o un'intestazione al suo inizio, e (meno comunemente) una chiusura quale 'Fine del capitolo 1'.

I seguenti elementi possono essere usati per trascriverli:

- <head> contiene ogni tipo di intitolazione, per esempio, il titolo di una sezione, oppure l'intestazione di una lista o di un glossario.
- <trailer> contiene una formula di chiusura o un elemento a piè di pagina che compare alla fine di una divisione di un testo.

Esempio:

```
<div1 id='UNC1' n='Libro primo' type='Sezione'>
  <div2 id='UNC11' n='1' type='Capitolo'>
    <head>I. Mia moglie e il mio naso</head>
    <p>&mdash; Che fai? &mdash; mia moglie mi domandò,
vedendomi...</p>
```

(da *Uno, nessuno e centomila* di Pirandello)

Prosa, versi, testi drammatici



PROSA

I paragrafi che costituiscono una divisione testuale dovrebbero essere marcati con `<p>`

Esempio:

```
<body>
```

```
...
```

```
<p>A seguitare, non ci saremmo mai intesi; perché  
se a me stava a cuore la tigre, a lei il  
cacciatore. </p>
```

```
...
```

```
</body>
```

(Il brano è tratto da *Quaderni di Serafino Gubbio operatore* di Luigi Pirandello)

Versi



- `</>` contiene una singola linea, anche incompleta, di un verso. Tra gli attributi:
 - *part*: specifica se la linea è metricamente completa o meno. I valori legali sono: *F* per la parte finale di una linea incompleta, *Y* se la linea è metricamente incompleta, *N* se la linea è completa, o se nessuna assunzione è fatta sulla sua completezza, *I* per la parte iniziale di una linea incompleta, *M* per la parte mediana di una linea incompleta.
- `<lg>` contiene un gruppo di versi che costituiscono un'unità formale, per esempio una stanza, un refrain, un paragrafo in versi, etc.

Esempio di codifica del verso



Qui, per esempio, abbiamo l'inizio di un testo poetico del quale sono marcate le linee di verso e le stanze:

```
<lg n='I'>
  <l>Spesso il male di vivere ho incontrato:</l>
  <l>era il rivo strozzato che gorgoglia,</l>
  <l>era l'incartocciarsi della foglia</l>
  <l>riarsa, era il cavallo stramazzato.</l>
</lg>
```

(Il brano è tratto da *Ossi di seppia* di Eugenio Montale)

Si noti come l'elemento `</>` marchi il verso, non le linee tipografiche: l'originaria lineazione delle poche linee precedenti non è stata resa esplicita da questa codifica, e potrebbe andare perduta. L'elemento `<lb/>` descritto nella sezione *Numeri di pagina e di linea*, potrebbe essere usato, se lo si desidera, per marcare le linee tipografiche.

Testi drammatici



- `<sp>` contiene una singola battuta in un testo drammatico, o un brano presentato come tale (cioè per la recitazione) in un testo in prosa o in versi. Tra gli attributi:
 - `who`: identifica mediante un ID il soggetto parlante
- `<speaker>` contiene una forma particolare di intestazione o etichetta, che fornisce i nomi di uno o più parlanti in un testo o frammento drammatico.
- `<stage>` contiene qualsiasi tipo di didascalie e direttive di scena all'interno di un testo o di un frammento di testo drammatico. Tra gli attributi:
 - `type`: indica il tipo di direttiva di scena. I valori consigliati includono *ENTRATA*, *USCITA*, *MESSA IN SCENA*, *DIZIONE*, etc.

Esempio di versi in testi teatrali



Qualche volta, particolarmente nei testi teatrali, le linee di verso sono suddivise tra più parlanti. Il modo più semplice per codificare questo fenomeno è usare l'attributo *part* per indicare che le linee così frammentate sono incomplete, come in questo esempio (ripreso dal primo atto, scena prima, dell'*Amleto* di W. Shakespeare):

```
<div1 type='Atto' n='I'><head>ATTO I</head>
  <div2 type='Scena' n='1'><head>SCENA 1</head>
    <stage rend='italic'>Francesco di scolta. Entra
    Bernardo</stage>
    <sp><speaker>Bern<l part='y'>Chi è là?
    <sp><speaker>Fran<l>Rispondi tu, prima, e manifestati.
    <sp><speaker>Bern<l part='i'>Viva il re!
    <sp><speaker>Fran<l part='m'>Bernardo?
    <sp><speaker>Bern<l part='f'>Quello.
    <sp><speaker>Fran<l>Esattamente venite alla vostr'ora.
    ...
  </div2>
..
</div1>
```

Esempio di dialogo in testo teatrale



L'esempio (tratto da *Quaderni di Serafino Gubbio operatore* di Luigi Pirandello) mostra come dovrebbe essere codificato il dialogo presentato in un'opera in prosa come se fosse un dialogo teatrale. Esso mostra anche l'uso dell'attributo *who* per veicolare un codice che identifichi chi recita un determinato dialogo:

```
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>
<p>&mdash; Che vuole che le dica? Lei in questo momento, lo
riconoscerà è molto eccitato.
<sp who='FER'><speaker>Carlo Ferro</speaker>
<p>&mdash; Ma posso esser calmo?
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>
<p>&mdash; Ah, capisco...
<sp who='FER'><speaker>Carlo Ferro</speaker>
<p>&mdash; Ne ho ragione, mi sembra!
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>
<p>&mdash; Sì, senza dubbio! Ma in tale stato, caro Ferro,
è anche molto facile esagerare.</sp>
```

Numeri di pagina e di linea



Le interruzioni di pagina e di linea possono essere marcate con i seguenti elementi vuoti.

- `<pb/>` segnala i confini tra una pagina di un testo e la successiva in un sistema di riferimento standard.
- `<lb/>` segnala l'inizio di una nuova linea (tipografica) in una certa edizione o versione di un testo.

Questi elementi marcano un punto singolo nel testo, non una sua porzione.

L'attributo globale *n* dovrebbe essere usato per fornire il numero della pagina o di linea che inizia con il marcatore.

Inoltre, questi due elementi condividono il seguente attributo:

ed

indica l'edizione o versione in cui l'interruzione della pagina è localizzata a questo punto.

Specificazioni nell'uso degli attributi



- Quando si lavora da un originale impaginato, è spesso utile registrare la sua impaginazione, anche solo per semplificare una successiva correzione delle bozze. Memorizzare l'interruzione delle linee potrebbe essere utile per la stessa ragione; il trattamento della divisione sillabica di parole a fine linea nell'originale a stampa richiederà una certa considerazione.
- Qualora si intenda marcare l'impaginazione, etc., di più di un'edizione, occorre specificare l'edizione in questione usando l'attributo *ed*, e fornire tutti i marcatori necessari. Per es. nel passaggio seguente (tratto da *Il tempo ritrovato* di Marcel Proust) indichiamo dove cade l'interruzione di pagina in due edizioni differenti (*ED1* e *ED2*).

Esempio



```
<p>La notte era bella come nel 1914, e Parigi
  altrettanto minacciata. Il chiaro di luna pareva
  quasi un vago, ininterrotto magnesio che
  permettesse di fissare un'ultima volta le immagini
  notturne di certe belle vedute, come place
<pb ed='ED1' n='128' />
Vend&ocirc;me o place de la Concorde, alle quali il
  mio terrore per le bombe che forse le avrebbero
  fra poco distrutte,
<pb ed='ED2' n='97' /> dava per contrasto, nella loro
  bellezza ancora intatta, una sorta di
  pienezza[...]</p>
```

Elementi 'pietra miliare'



Gli elementi `<pb/>` e `<lb/>` sono casi particolari della classe di cosiddetti elementi 'pietra miliare', che marcano i punti di riferimento all'interno di un testo.

La *TEI Lite* prevede anche un elemento generico `<milestone>` che non è ristretto a casi particolari ma può segnalare ogni tipo di punto di riferimento: per esempio, un'interruzione di colonna, l'inizio di un nuovo tipo di sezione non marcata altrimenti, etc.

Questo elemento ha i seguenti attributi e descrizione:

`<milestone>`

segnala il confine tra le sezioni di un testo, indicato dai cambiamenti in un sistema di riferimento standard. Tra gli attributi:

ed: indica l'edizione o versione a cui si riferisce il riferimento.

unit: indica quale tipo di sezione sta cambiando quando si incontra questo riferimento.

I nomi usati per i tipi di unità e per le edizioni cui ci si riferisce con gli attributi *ed* e *unit* possono essere scelti liberamente, ma devono essere documentati nella testata.

Si può usare l'elemento `<milestone>` per rimpiazzare gli altri, o usare questi nel loro insieme; i due metodi non andrebbero mescolati arbitrariamente.

Codifica di espressioni evidenziate



1. Cambio di stile nel carattere
2. Citazioni e caratteristiche correlate
3. Parole in lingua straniera

1. Cambio di stile del carattere



- Le parole o le espressioni evidenziate sono quelle rese visualmente differenti dal resto del testo, generalmente per mezzo di un cambio nella fonte tipografica, nello stile di scrittura o nel colore dell'inchiostro, al fine di richiamare l'attenzione del lettore.
- L'attributo globale *rend* può essere associato ad ogni elemento e usato ovunque sia necessario specificare i dettagli dell'evidenziazione per esso usata. Per esempio un titolo reso in grassetto può essere marcato *rend='Bold'*, uno in corsivo *rend='Italic'*.

L'elemento di evidenziazione generica



Non è sempre possibile o desiderabile interpretare le ragioni di tali cambiamenti di aspetto in un testo. In tali casi, l'elemento `<hi>` può essere usato per marcare una sequenza di testo evidenziato senza fare alcuna assunzione circa il suo status.

`<hi>` marca una parola o frase in quanto graficamente distinta dal testo circostante, senza che si faccia alcuna dichiarazione sulle ragioni dell'evidenziazione.

Nel seguente esempio (tratto da *Quaderni di Serafino Gubbio operatore* di L. Pirandello) l'uso dei vari tipi di carattere è registrato ma non interpretato:

```
<p>&laquo;Siete proprio necessario voi? Che cosa siete voi? <hi rend='italic'>Una mano che gira la manovella</hi>. Non si potrebbe fare a meno di questa mano?[...]</p>
```

Ulteriori elementi per l'evidenziazione



- `<emph>` marca parole o espressioni che sono messe in risalto o enfatizzate per un effetto linguistico o retorico.
- `<foreign>` identifica una parola o un'espressione appartenente a una lingua diversa da quella del testo circostante.
- `<mentioned>` marca parole o espressioni citate o riportate.
- `<term>` contiene una o più parole o designazioni simboliche che vengono considerate come termini tecnici.

L'elemento di codifica dei titoli



`<title>` contiene il titolo di un'opera, sia essa articolo, libro, giornale, o collana ogni titolo alternativo o sottotitolo. Tra gli attributi:

level: indica se questo è il titolo di un articolo, libro, giornale, collana o materiale inedito. I valori permessi sono: *M* per titoli monografici (libro, collezione o altro oggetto pubblicato come singolo oggetto includendo opere singole o lavori in più volumi); *S* (titolo di una collana); *J* (titolo di giornale); *U* per titoli di materiale inedito (includere tesi e dissertazioni non ancora pubblicate da un editore commerciale); *A* per titolo analitico (articolo, poema, o altro materiale pubblicato come parte di un oggetto più grande).

type: classifica il titolo secondo un'adeguata tipologia. I valori possibili includono: *ABBREVIATO*; *PRINCIPALE*; *SUBORDINATO* (per sottotitoli o titoli di sezione); *PARALLELO* (per titoli alternativi, spesso in un'altra lingua, con cui l'opera è conosciuta).

Altri elementi di evidenziazione



Alcune caratteristiche (specificamente citazioni e commenti) possono trovarsi in un testo evidenziate oppure incluse tra virgolette.

In entrambi i casi dovrebbero essere usati gli elementi `<q>` e `<gloss>` (come discusso nella sezione seguente).

Qualora occorra registrare la resa tipografica, si usi l'attributo globale *rend*.

2. Citazioni e caratteristiche correlate



Come i cambi di carattere tipografico, le virgolette sono convenzionalmente usate per contraddistinguere diverse caratteristiche all'interno di un testo, tra le quali la più frequente è la citazione. Quando è possibile, è consigliabile codificare la caratteristica strutturale sottostante, piuttosto che segnalare il semplice fatto che una citazione appare nel testo, usando i seguenti elementi:

<q> contiene una citazione, manifesta o meno - una rappresentazione di discorso o pensiero marcata come se fosse espressa da qualcun altro (sia essa realmente citata o meno); in narrativa, le parole sono di solito quelle di un personaggio o di chi parla; nei dizionari, <q> può essere usato per indicare esempi, reali o inventati, dell'uso di un termine.

Tra gli attributi:

type: può essere usato per indicare se il brano citato è parlato o pensato, oppure per caratterizzarlo più finemente. Valori esemplificativi sono: *DIRETTO* (per la rappresentazione del discorso diretto, di solito segnalato dalle virgolette); *PENSATO* (per la rappresentazione del pensiero, per esempio il monologo interiore).

who: identifica colui che pronuncia il discorso diretto.

Altri elementi per citazioni e correlati



- <mentioned> marca parole o espressioni citate o riportate.
- <soCalled> contiene una parola o espressione per cui l'autore o il narratore non si assume la responsabilità, segnalate ad esempio tramite l'uso di virgolette o corsivi.
- <gloss> marca una parola o espressione che fornisce una glossa o definizione per qualche altra parola o espressione. Tra gli attributi:
 - *target*: identifica la frase o parola associata.

Esempio di uso di <q>



Le citazioni possono essere accompagnate dal riferimento alla fonte o a colui che parla, usando l'attributo *who*, sia che la fonte venga esplicitamente segnalata o meno nel testo, come nel seguente esempio (tratto da *Suo marito* di L. Pirandello):

```
<p>&mdash; <q who='Giustino'>Hai veduto? hai veduto?</q>
&mdash; le singhiozzava intanto Giustino &mdash; <q>Se
n'è andato,.. Rirì se n'è andato, perché noi non
c'eravamo... tu non c'eri... e neanche io c'ero più... e
allora il povero piccino ha detto: &laquo;<q who='Rirì'>E
che ci faccio più io qua?</q> &raquo; e se n'è andato...
Se ti vedesse qua ora... Vieni! vieni! Se ti vedesse
qua...</q>
</p>
```

Questo esempio mostra anche come le citazioni possano trovarsi all'interno di altre citazioni: un personaggio (Giustino) ne cita un altro (Rirì).

3. Parole in lingua straniera



Parole o espressioni che non sono nella lingua principale del testo, possono essere marcate come tali in due modi.

Se la parola o l'espressione è già marcata per qualche ragione, l'elemento indicato dovrebbe avere un valore per l'attributo globale *lang*, indicante la lingua usata.

Se non c'è alcun elemento applicabile, può essere usato l'elemento *<foreign>*, sempre corredato dell'attributo *lang*.

Codifica della lingua come attributo



- Giovanni ha molto `<foreign lang='fra'>savoir-faire</foreign>`.
- Hai letto `<title lang='ted'>Die Dreigroschenoper</title>?`
- `<mentioned lang='fra'>Savoir-faire</mentioned>` è il termine francese per competenza.
- La corte ha emanato un mandato di `<term lang='lat'>mandamus</term>`.

Come mostrano questi esempi, l'elemento `<foreign>` non dovrebbe essere usato per marcare parole straniere se altri elementi più specifici come `<title>`, `<mentioned>`, o `<term>` possono essere applicati. L'attributo globale `lang` può essere aggiunto a qualunque elemento per indicare che usa una lingua diversa da quella del testo circostante.

Codifica delle note



`<note>` contiene una nota o annotazione. Tra gli attributi:

- `type`: descrive il tipo di nota.
- `resp`: indica chi è il responsabile dell'annotazione. Il valore potrebbe essere: *AUTORE*, *EDITORE*, *TRADUTTORE*, etc. o le iniziali di colui che ha inserito le annotazioni.
- `place`: indica dove appare la nota nel testo originale. Esempi di valori sono: *INTRALINEA*, *INTERLINEA*, *SINISTRA*, *DESTRA*, *PIEDE PAGINA*, *FINE*, per le note che appaiono rispettivamente come paragrafi marcati nel corpo del testo, tra le righe del testo, al margine sinistro o destro, a piè di pagina, oppure alla fine del capitolo o del volume.
- `target`: indica il punto di aggancio della nota, o l'inizio del brano a cui si riferisce la nota.
- `targetEnd`: indica la fine del brano a cui si riferisce la nota, se la nota non è inserita nel testo a quel punto.
- `anchored`: indica se il testo originale mostra l'esatto punto di riferimento per la nota.

Alcune precisazioni



- Se possibile, il corpo di una nota dovrebbe essere inserito nel testo nel punto esatto in cui appare il relativo rimando o riferimento. Ciò non è possibile, per esempio, con i *marginalia*, che potrebbero non essere collegati ad un luogo preciso. Per semplicità, può essere adeguato posizionare le note marginali prima del paragrafo rilevante o di altri elementi. Le note possono anche essere posizionate in una sezione separata del testo (come avviene per le note a fine testo nei libri a stampa) e collegate al testo mediante i rispettivi attributi *target*.
- L'attributo *n* può essere usato per fornire il numero o l'identificatore di una nota se ciò è richiesto. Per distinguere tra le note d'autore e le note editoriali (qualora siano presenti entrambe), deve essere usato l'attributo *resp*; altrimenti occorre dichiarare esplicitamente di che tipo siano nella testata TEI.

Esempi di codifica delle note



Le collezioni sono insiemi di distinte entità o oggetti di ogni specie.

```
<note place='foot' n='1'> Spieghiamo sotto perché usiamo l'insolito termine <mentioned>collezione</mentioned> al posto del più usuale <mentioned>serie</mentioned>. Il nostro utilizzo corrisponde all'<mentioned>aggregato</mentioned> dei molti scritti matematici e al senso dell'<mentioned>ordine</mentioned> trovato in scritti logici più vecchi.
</note>
```

```
<lg id='RAM609'>
<note place='margin'>Il corso è finalmente espletato</note>
<l>e ora questo incantesimo è stato rotto: ancora una volta</l>
<l>ho visto l'oceano verde,</l>
<l>e ho guardato avanti lontano, e ho visto le cose piccole</l>
<l>come mai gli altri le avrebbero viste &dash;</l>
</lg>
```

Riferimenti incrociati e collegamenti



1. I riferimenti incrociati espliciti o i collegamenti da un punto di un testo ad un altro nello stesso documento possono essere codificati usando gli elementi descritti nella sezione *Riferimenti incrociati semplici*.
2. Riferimenti o collegamenti ad elementi di qualche altro documento XML, o di parti di documenti non-XML, possono essere codificati usando i puntatori estesi TEI descritti nella sezione *Puntatori estesi*.
3. I legami impliciti (come le associazioni tra due testi paralleli, o tra un testo e la sua interpretazione) possono essere codificati usando gli attributi di collegamento discussi nella sezione *Attributi di collegamento*.

1. Riferimenti incrociati semplici



Un riferimento incrociato da un punto a un altro all'interno di un singolo documento può essere codificato usando uno dei seguenti elementi:

`<ref>` un riferimento a un'altra posizione nel documento corrente, in termini di uno o più elementi identificabili, eventualmente caratterizzato da testo supplementare o da un commento.

`<ptr>` un puntatore ad un'altra posizione nel documento corrente, in termini di uno o più elementi identificabili.

Attributi per riferimenti incrociati semplici



Questi elementi condividono i seguenti attributi:

- *target*: specifica la destinazione di un puntatore come uno o più identificatori XML.
- *type*: categorizza il puntatore per qualche aspetto, usando qualsiasi conveniente insieme di categorie.
- *targType*: specifica il tipo (o i tipi) di elemento al quale questo puntatore può rinviare.
- *crDate*: specifica quando il puntatore è stato creato.
- *resp*: specifica il creatore del puntatore.

Differenza <ptr> e <ref>



La differenza tra questi due elementi è che <ptr> è un elemento vuoto, che individua semplicemente un punto dal quale si deve fare un collegamento, mentre <ref> può anche contenere del testo - solitamente il testo stesso del riferimento incrociato. L'elemento <ptr> invece sarà utilizzato per un riferimento incrociato rappresentato da dispositivi non verbali come un simbolo o un'icona, oppure, in un testo elettronico, da un bottone. È anche utile nei sistemi di produzione di documenti, dove un programma di formattazione può generare la corretta forma verbale del riferimento incrociato.

Le seguenti due forme, per esempio, sono logicamente equivalenti (assumendo che noi abbiamo documentato altrove l'esatta forma verbale dei riferimenti rappresentati dagli elementi <ptr>):

```
Vedere specialmente <ref target='SEZ12'>sezione 12 a  
pagina 34</ref>. Vedere specialmente <ptr  
target='SEZ12' />.
```

Valore dell'attributo *target*



Il valore di un attributo *target* deve essere un identificatore XML nel documento XML corrente. Questo implica che il passaggio o l'espressione a cui si punta deve avere un identificatore, e deve perciò essere codificato come un elemento di qualche tipo.

Nel seguente esempio, il riferimento incrociato punta ad un elemento `<div1>`:

```
...  
vedere specialmente <ptr target='SEZ12'/>  
... <div1 id='SEZ12'><head>Riguardo  
l'identificazione...
```

Attributo *id*



Poiché l'attributo *id* è globale, qualsiasi elemento nel documento può essere indicato in questo modo.

Nel seguente esempio, ad un paragrafo è stato dato un identificatore cosicché possa essere 'puntato':

```
...  
questo è discusso nel <ref target='pspec'>  
paragrafo sui collegamenti</ref>  
...  
<p id='pspec'>I collegamenti possono essere  
costruiti per qualsiasi tipo di elemento  
...
```

Attributo *targType*



L'attributo *targType* può essere usato per specificare che l'elemento puntato deve essere di un particolare tipo, come nel seguente esempio:

...

```
questo è discusso nella <ref target=dspec  
  targType='div1 div2'> sezione sui  
  collegamenti</ref>
```

Questo riferimento dovrebbe dare errore se l'elemento con identificatore *dspec* non fosse una *<div1>* o una *<div2>*. Si noti comunque che questo controllo non può essere effettuato da un *parser XML*, poiché esso può solo controllare che esista qualche elemento identificato come *dspec*.

Altri attributi



L'attributo *type* può essere usato per categorizzare il legame rappresentato dall'indicatore in ogni modo appropriato. Gli attributi *resp* e il *crDate* possono anche essere usati per rappresentare la persona o l'entità responsabile della creazione del collegamento, e la sua data di creazione, come nell'esempio seguente:

...

```
questo è discusso nella  
<ref type='xref' resp='auto' crdate='210595'  
  target='dspec' targtype='div1 div2'>  
sezione di collegamento</ref>
```

Questi attributi saranno verosimilmente utilizzati nei sistemi ipertestuali che contengono moltissimi puntatori, usati per vari scopi e creati con vari mezzi.

Altri elementi



Talvolta la destinazione di un riferimento incrociato non corrisponde a nessuna particolare caratteristica di un testo, e non può dunque essere codificata come elemento di un qualche tipo. Se la destinazione desiderata è semplicemente un punto del documento corrente, il modo più semplice per marcarla è l'introduzione di un elemento `<anchor>` nel punto appropriato. Se il *target* è una sequenza di parole non altrimenti codificate, può essere introdotto l'elemento `<seg>` per marcarle. Questi due elementi sono descritti come segue:

- `<anchor>` specifica una posizione o un luogo all'interno di un documento affinché possa essere 'puntato'.
- `<seg>` identifica un frammento o un segmento di testo all'interno di un documento affinché possa essere 'puntato'. Tra gli attributi:
 - *type*: categorizza il segmento.

Uso di `<ref>`



Nel seguente esempio, degli elementi `<ref>` sono stati usati per rappresentare dei punti nel testo che sono collegati, in qualche modo, ad altre parti dello stesso testo; nel primo caso ad un luogo particolare, e nel secondo, ad una sequenza di parole:

```
Ritornando al <ref target='ABCD'>punto dove ero  
rimasto</ref>, ho rilevato che <ref  
target='EFGH'>tre parole</ref>, sono state  
sottolineate in rosso da un precedente lettore.
```



Uso di <anchor> e <seg>

Questa codifica richiede che gli elementi con gli identificatori specificati (*ABCD* e *EFGH* in questo esempio) si trovino in qualche punto dello stesso documento. Qualora non esista alcun elemento a cui attribuire tali identificatori, si possono usare gli elementi <anchor> e <seg>:

```
....<anchor type='segnalibro' id='ABCD' />....  
....<seg type='target' id='EFGH'>...</seg>...
```

L'attributo *type* dovrebbe essere usato (come sopra) per distinguere i diversi ruoli che questi elementi generici possono assumere in un testo. Altre applicazioni saranno discusse nella sezione *Attributi di collegamento*.



2. Puntatori estesi

Gli elementi <ptr> e <ref> possono essere usati solo per i riferimenti incrociati o i collegamenti le cui destinazioni si trovino nello stesso documento XML della loro origine. Inoltre, essi possono riferirsi esclusivamente ad elementi XML.

Gli elementi discussi in questa sezione non sono sottoposti a tali limitazioni.

<xptr> definisce un puntatore a un'altra posizione nel documento corrente o in un documento esterno.

<xref> definisce un puntatore ad un'altra posizione nel documento corrente o in un documento esterno, eventualmente caratterizzato da un testo supplementare o da un commento.

Attributi per puntatori estesi



In aggiunta agli attributi applicabili ai puntatori già discussi sopra nella sezione *Riferimenti incrociati semplici*, questi elementi condividono i seguenti attributi supplementari, che sono usati per specificare la destinazione di un riferimento incrociato o di un collegamento al posto dell'attributo *target*:

- *doc*: indica il documento all'interno del quale deve essere trovata il luogo richiesto; per *default* è il documento corrente.
- *from*: indica l'inizio della destinazione del puntatore sotto forma di un'espressione nella sintassi per i puntatori estesi TEI; per *default* è l'intero documento indicato dall'attributo *doc*.
- *to*: indica il punto finale della destinazione del puntatore come un'espressione nella sintassi puntatori estesi TEI; può essere specificato solo se è stato indicato l'attributo *from*.

Ancora sui puntatori estesi



Una specificazione completa del linguaggio formale usato per esprimere le destinazioni dei puntatori estesi TEI è troppo complessa per essere tratta in questa sede; qui ci limiteremo ad elencare solo alcune delle sue caratteristiche in genere più utili. La versione completa delle *Norme* dovrebbe essere consultata per maggiori dettagli.

Uso dell'attributo *doc*



Un elemento `<xptr>` (o `<xref>`) può riferirsi ad un intero documento esterno semplicemente fornendo un nome di entità XML come valore dell'attributo *doc*, come in questo esempio:

```
vedere <xref doc='P3'> la TEI  
Guidelines, in vari luoghi</xref>
```

Questo esempio ipotizza che sia stata dichiarata un'entità di sistema o pubblica con il nome *P3*. Questa dichiarazione può essere posta all'interno del file per l'estensione *litemods.ent* o in qualche altra maniera specifica al software in uso (come discusso nella sezione *Immagini e grafica*).

Sintassi per puntatori estesi



L'attributo *from* è usato per indicare una data posizione all'interno di qualunque documento specificato dall'attributo *doc*.

La specificazione si basa su un linguaggio formale, denominato 'sintassi per i puntatori estesi TEI', del quale qui sono forniti solo alcuni dettagli.

In questo linguaggio, le posizioni sono definite come una serie di *passi* (*steps*), ognuno dei quali identifica una certa parte del documento, spesso nei termini delle posizioni identificate dal passo precedente.

Per esempio potreste puntare alla terza frase del secondo paragrafo del capitolo secondo, selezionando il capitolo due nel primo passo, il secondo paragrafo nel secondo passo, e la terza frase nell'ultimo passo.

Un passo può essere definito in termini di concetti XML (come *parent*, *descendent*, *preceding*, etc.) o, più vagamente, in termini di modelli (*patterns*) di testo, di posizione delle parole o dei caratteri tipografici.

È altresì possibile usare una notazione estranea (non-XML), o specificare una posizione all'interno di un'immagine nei termini del suo sistema di coordinate.

Uso dell'attributo *from to*



Gli attributi *from* e *to* usano la stessa notazione.

Ciascuno 'punta' a qualche porzione del documento di destinazione; il puntatore esteso nella sua totalità 'punta' alla sezione che inizia con l'attributo *from* e continua fino al *to*.

Il primo passo in un percorso di posizionamento sarà spesso quello di specificare l'identificatore di qualche elemento all'interno del documento di destinazione, come in questo esempio:

```
<xptr doc='P3' from='id(SA)'\>
```

Parole chiave per puntatori estesi



Questo puntatore seleziona l'intero elemento che porta l'identificatore *SA* all'interno della entità *P3*. Se è richiesta una destinazione più specifica, possono succedersi altri passi. Le seguenti parole chiave sono disponibili per selezionare altri elementi nei termini delle loro relazioni l'elemento identificato dal passo precedente:

- *child*: elementi da questo contenuti.
- *ancestor*: elementi che lo contengono, direttamente o indirettamente
- *previous*: elementi con lo stesso genitore di questo, ma che lo precedono nel documento.
- *next*: elementi con lo stesso genitore di questo e che lo seguono nel documento.
- *preceding*: elementi nel documento che iniziano prima di questo, senza tenere conto dei loro genitori.
- *following*: elementi nel documento che iniziano dopo di questo, senza tenere conto dei loro genitori.

Altri dettagli sui puntatori estesi



- Ognuna di queste parole chiave implica un particolare insieme di elementi (l'insieme dei figli, l'insieme dei genitori, l'insieme dei discendenti, etc.); per specificare a quale elemento dell'insieme ci stiamo riferendo, la parola chiave può essere seguita da una lista tra parentesi contenente:
- un numero positivo o negativo, che indichi quale dei molti possibili elementi trovati sia quello designato (+1 per indicare il primo elemento incontrato, partendo dalla posizione corrente, e -1 per indicare l'ultimo), o la parola chiave *ALL*, per indicare che ci si deve riferire a tutti gli elementi dell'insieme;
- un identificatore generico, indicante il tipo dell'elemento richiesto, o un asterisco (*) indicante che qualsiasi tipo di elemento andrà bene;
- una serie di nomi e valori di attributo, indicante che l'elemento selezionato deve avere attributi con i nomi e i valori specificati, se ve ne è qualcuno.

Esempio



Continuando l'esempio precedente, il riferimento seguente selezionerà il terzo elemento `<p>` direttamente contenuto da qualsivoglia elemento recante l'identificatore SA:

```
<xptr doc='P3' from='id (SA) child (3 p)'/>
```

Similmente, assumendo che l'entità *P3* è in effetti un riferimento alla versione XML delle *Norme TEI*, il puntatore seguente seleziona la sezione 14.2.2 di tale pubblicazione, dove (per combinazione) viene formalmente definita la sintassi del puntatore esteso:

```
Per ulteriori dettagli, vedere <ref doc='P3'  
from='id (SA) child (2 div2) child (2 div3)'  
> TEI estesa indicatore sintassi definita</ref>
```

Uso integrato di <from> e <to>



Normalmente, l'obiettivo di un riferimento incrociato sarà adeguatamente definito dall'attributo *from*. Per alcuni documenti, tuttavia, potrebbe essere più comodo definire sia l'obiettivo iniziale che quello finale. Come notato sopra, per questo fine viene fornito l'attributo *to*.

Per esempio:

```
<xptr doc='p1' from='id (xyz)' to='id (abc)'\/>
```

è un puntatore esteso la cui destinazione è la sequenza che inizia con qualsivoglia elemento recante l'identificatore *XYZ* nel documento *P1*, e che termina con qualsiasi elemento recante l'identificatore *ABC*, sempre, come ovvio, all'interno del medesimo documento. Sono tutti gli elementi che si trovano tra questi due, indipendentemente dalla loro struttura.

Esempi



Questa sintassi permette di costruire facilmente specificazioni assai complesse. Per esempio, il seguente riferimento selezionerà il più prossimo elemento <head> che abbia un attributo *lang* con valore *LAT*, e che si trovi prima dell'inizio dell'elemento con l'identificatore *SA*:

```
<xptr doc='P3' from='id(SA) preceding (1 head lang lat)'\/>
```

Se non viene fornito nessun valore per l'attributo *doc*, si assume che ci si riferisca al documento corrente. I riferimenti seguenti, dunque, sono semanticamente equivalenti. Entrambi indicano l'elemento con l'identificatore *X1* all'interno del documento corrente:

```
<ptr target='X1'\/>  
<xptr from='id (X1)'\/>
```

3. Attributi di collegamento



I seguenti attributi di collegamento speciali, sono definiti per ogni elemento nella DTD *TEI Lite*:

- *ana*: collega un elemento con la sua interpretazione.
- *corresp*: collega un elemento con uno o più altri elementi corrispondenti.
- *next*: collega un elemento con l'elemento seguente in un aggregato.
- *prev*: collega un elemento con l'elemento precedente in un aggregato.

Attributo *ana*



L'attributo *ana* (analisi) si usa nei casi in cui una serie di analisi astratte o interpretazioni siano state definite in qualche punto all'interno di un documento, come discusso estesamente nella sezione *Interpretazione ed analisi*.

Per esempio, un'analisi linguistica della frase 'Renzo ama Lucia' può essere codificata come segue:

```
<seg type='frase' ana='SVO'>  
  <seg type='lex' ana='NP1'>Renzo</seg>  
  <seg type='lex' ana='VV1'>ama</seg>  
  <seg type='lex' ana='NP1'>Lucia</seg>  
</seg>
```

Questa codifica implica l'esistenza da qualche altra parte nel documento di elementi con identificatori *SVO*, *NP1*, *VV1*, dove viene spiegato il significato di tali particolari codici. Si noti l'uso dell'elemento `<seg>` per marcare componenti particolari dell'analisi, distinti dall'attributo *type*.

Attributo *corresp*



L'attributo *corresp* (corrispondente) fornisce un modo semplice per rappresentare alcune forme di corrispondenza tra due elementi in un testo. Per esempio, in un testo multilingue, può essere usato per collegare traduzioni equivalenti, come nel caso seguente:

```
<seg lang='ITA' id='IT1' corresp='FR1'>John ama  
Nancy</seg>  
<seg lang='FRA' id='IT1' corresp='IT1'>Jean aime  
Nancy</seg>
```

Lo stesso meccanismo può essere usato per molti altri scopi. Nell'esempio seguente, viene usato per rappresentare delle corrispondenze anaforiche tra 'lo spettacolo' e 'Shirley', e tra 'NBC' e 'il network':

```
<p><title id='shirley'>Shirley</title>, concluso venerdì  
sera ha debuttato solo un mese fa, non è segnato nelle  
nuove schede della <name id='nbc'>NBC</name>, sebbene il  
<seg id='network' corresp='nbc'> network</seg> dica che  
lo <seg id='show' corresp='shirley'>show</seg> sia  
ancora considerato...
```

Attributi *nexte prev*



Gli attributi *next* e *prev* forniscono un modo semplice per legare insieme le componenti di un elemento discontinuo, come nell'esempio seguente:

```
<p>&mdash; <q id='Q1a' next='Q1b'>Ma sì!</q>  
&mdash; esclamai. &mdash;  
<q id='Q1b' prev='Q1a'>Ha la febbre, e  
forte. Aspetti. Chiamo il signor Cavalena.  
Il nostro padrone di casa è medico.</q>
```

Interventi editoriali



Il processo di codifica di un testo elettronico ha molto in comune con il processo di edizione di un manoscritto o di altri testi per la pubblicazione a stampa.

In entrambi i casi un editore coscienzioso può voler registrare sia lo stato originale della fonte, sia ogni correzione editoriale o altro cambiamento introdotto.

Gli elementi discussi in questa sezione e nella prossima forniscono alcune risorse adatte a tali necessità.

Elementi <corr> e <sic>



La seguente coppia di elementi può essere usata per marcare delle correzioni, cioè dei cambiamenti editoriali introdotti laddove l'editore ritenga che l'originale sia erroneo:

- **<corr>** contiene la forma corretta di un passaggio evidentemente erroneo nel testo. Tra gli attributi:
 - *sic*: dà la forma originale dell'errore apparente nel testo.
 - *resp*: indica l'editore o il trascrittore che ha proposto la correzione contenuta dall'elemento *<corr>*.
 - *cert*: indica il grado di certezza attribuito alla correzione contenuta dall'elemento *<corr>*.
- **<sic>** contiene un testo riprodotto anche se evidentemente non corretto o impreciso. Tra gli attributi:
 - *corr*: dà una correzione di un errore palese in un testo.
 - *resp*: indica l'editore o il traduttore responsabile per il suggerimento della correzione.
 - *cert*: indica il grado di certezza attribuito alla correzione.

Elementi <orig> e <reg>



La seguente coppia di elementi può essere usata per marcare la normalizzazione, cioè i cambiamenti editoriali introdotti per salvaguardare la coerenza di un testo o per modernizzarlo:

- <orig> contiene la forma originale di una lezione della quale è data una forma regolarizzata in un valore di un attributo. Tra gli attributi:
 - *reg*: dà una forma regolarizzata (normalizzata) di un testo.
 - *resp*: identifica il responsabile individuale della regolarizzazione della parola o frase.
- <reg> contiene una lezione che è stata regolarizzata o normalizzata in qualche modo. Tra gli attributi:
 - *orig*: dà la forma non regolarizzata del testo come si trova nella copia originale.
 - *resp*: identifica il responsabile individuale per la regolarizzazione della parola o frase.

Esempio



Per esempio il seguente passo da *Il Turno* (Cap. II) di Luigi Pirandello:

Don Diego non era ancora potuto entrare in casa della promessa sposa. Stellina minacciava di graffiargli la faccia, di cavargli tutti e due gli occhi, se egli si fosse arrischiato di presentarsi a lei. Il Ravi, s'intende, non parlava a don Diego di queste minacce della figliuola; diceva soltanto che bisognava avere un po' di piazienza, perché le ragazze, oh Dio, si sa...

tratto dall'edizione Quattrini del 1920 introduce l'errore 'piazienza' per 'pazienza', come rileva Costanzo, e contiene la forma arcaica 'figliuola'. Gli emendamenti possono essere così codificati:

```
<p>Don Diego non era ancora potuto entrare in casa della promessa  
sposa. Stellina minacciava di graffiargli la faccia, di cavargli  
tutti e due gli occhi, se egli si fosse arrischiato di presentarsi  
a lei. Il Ravi, s'intende, non parlava a don Diego di queste  
minacce della <reg orig='figliuola' resp='COD'>figliuola</reg>;  
diceva soltanto che bisognava avere un po' di <corr sic='piazienza'  
resp='Costanzo'>pazienza</corr>, perché le ragazze, oh Dio, si sa...  
</p>
```

Omissioni, soppressioni e aggiunte



In aggiunta alle correzioni o normalizzazioni di parole o frasi, gli editori e i trascrittori possono anche fornire degli elementi mancanti, oppure trascrivere degli elementi che erano stati cancellati o depennati con una riga nel testo originale.

Inoltre, alcuni passaggi possono essere particolarmente difficili da trascrivere perché è difficile decifrarli sulla pagina originale.

Elementi per la codifica dei fenomeni (aggiunte e omissioni)



- `<add>` contiene lettere, parole o frasi inserite in un testo da un autore, *scriba*, commentatore o correttore. Tra gli attributi:
 - *place*: se l'aggiunta è iscritta nel testo, indica la collocazione del testo supplementare. I valori ad esempio possono essere *INTRALINEA* (sulla stessa riga), *SOPRALINEA* (sopra la riga), *INTERLINEA*, *SINISTRA* (nel margine di sinistra), *DESTRA* (nel margine di destra), *SOPRA*, *SOTTO*.
- `<gap>` indica un punto dove si è omesso del testo in una trascrizione, sia per ragioni editoriali descritte nella testata TEI (nella sezione sui criteri di selezione), sia perché il materiale è illeggibile, o impercettibile. Tra gli attributi:
 - *desc*: dà una descrizione del testo omesso.
 - *resp*: indica l'editore, il traduttore o il codificatore responsabile per la decisione di non fornire nessuna trascrizione del testo, da cui l'applicazione del marcatore `<gap>`.

Elementi per la codifica dei fenomeni (cancellazioni e soppressioni)



- `` contiene una lettera, parola o passaggio cancellati, marcati come cancellati, o altrimenti indicati come superflui o spuri, nella copia originale da parte di un autore, scriba, commentatore o correttore. Tra gli attributi
 - *type*: classifica il tipo di cancellazione usando una tipologia adeguata.
 - *status*: può essere usato per indicare cancellature errate, per esempio tratti di linea che coprono troppo o troppo poco testo.
 - *hand*: significa la mano di chi ha eseguito la cancellatura.
- `<unclear>` contiene una parola, una frase o un brano che non può essere trascritto con certezza perché è illeggibile o impercettibile nell'originale. Tra gli attributi:
 - *reason*: indica la ragione per la quale è difficile trascrivere il materiale.
 - *resp*: indica il responsabile individuale per la trascrizione della lettera, delle parole o del passaggio contenuti nell'elemento `<unclear>`.

Esempio di cancellazione



Questi elementi possono essere usati per registrare cambiamenti fatti da un editore o da un trascrittore o, nel caso di materiale manoscritto, da un autore o dal copista. Per esempio, se la fonte originale per un testo elettronico riporta:

I seguenti elementi sono forniti per per semplici interventi editoriali

Può sembrare opportuno correggere quest'errore palese, ma allo stesso tempo registrare le soppressioni del secondo 'per', superfluo. Così:

I seguenti elementi sono forniti per
`<del hand='LB'>per` semplici interventi editoriali

Il valore *LB* nell'attributo *hand* indica che 'LB' ha corretto la duplicazione della preposizione 'per'.

Esempio di aggiunta e cancellazione



Se l'originale riporta:

I seguenti elementi forniti per per
semplici interventi editoriali

(cioè, se il verbo fosse stato inavvertitamente omesso)
il testo corretto potrebbe essere il seguente:

I seguenti elementi <add
hand='LB'>sono</add> forniti per
<del hand='LB'>per semplici
interventi editoriali

Il valore *LB* nell'attributo *hand* indica che 'LB' ha
corretto l'omissione del verbo e la duplicazione della
preposizione *per*.

Esempio di intervento sulla variantistica d'autore



Questi elementi non sono limitati a cambiamenti
introdotti da un editore; essi possono anche essere
usati per registrare varianti d'autore nel manoscritto.

In un manoscritto de *I limoni* di Eugenio Montale
l'autore ha prima di tutto scritto 'tra gli alberi dai
nomi poco usati', poi ha tolto le parole 'gli alberi' e
ha inserito le parole 'le piante'; questo potrebbe
essere codificato come segue:

```
<l>tra <del hand='Montale'  
  type='overstrike'>gli alberi</del>  
<add hand='Montale' place='supralinear'>le  
  piante</add> dai nomi poco usati</l>:
```

Elementi <unclear> e <gap>



Similmente, gli elementi <unclear> e <gap>, possono essere usati insieme per indicare le omissioni di materiale illeggibile; l'esempio seguente (che riporta un passo dal *Convivio* di Dante in un'edizione a cura di C. Vasoli e D. De Robertis) mostra inoltre l'uso di <add> per una *emendatio* congetturale:

```
[...]che avessero sì leggiere le
  <unclear> <gap reason='indecifrabile'>
</unclear>
<add hand='ed'>non</add> fittizie parole
  apprese; [...]
```

Rapporto e <gap>



L'elemento identifica un passaggio che è stato trascritto nel testo elettronico nonostante fosse stato soppresso sull'originale, mentre <gap> indica la posizione di un passaggio che è stato omesso dal testo elettronico indipendentemente dal fatto che fosse leggibile o meno. Un corpus linguistico scritto in una data lingua, ad esempio, può omettere lunghe citazioni in lingue straniere:

```
<p>...Un esempio di una lista che compare nel libro
mastro del feudo di <name type='luogo'>
Koldinghus</name> <date>12/1611</date> è fornito a
piè di pagina. Esso mostra la riscossione degli
introiti di una rendita da parte di una vendita di
miele.</p>
<q><gap desc='citazione dal libro mastro' reason='in
danese'></q>
<p>Una descrizione della totale struttura dei conteggi
è ancora una volta...
```

Nomi, date, numeri e abbreviazioni



Lo schema di codifica TEI definisce elementi per un grande numero di dati 'strutturati' che possono apparire quasi ovunque in ogni tipo di testo.

Questi tipi di dati possono essere di particolare interesse per varie discipline; essi si riferiscono in generale ad oggetti esterni al testo stesso (come i nomi di persona e di luogo, i numeri e le date).

Essi inoltre pongono particolari problemi in molte applicazioni di trattamento automatico del linguaggio naturale (NLP), a causa della molteplicità di forme sotto cui possono essere presentate all'interno di un testo.

Gli elementi qui descritti riducono la difficoltà di sottoporre a trattamento informatico i testi che contengono tali informazioni, rendendole esplicite.

1. Nomi
2. Date
3. Numeri
4. Abbreviazioni

1. Nomi ed espressioni referenziali



Una *espressione referenziale* è un'espressione che si riferisce ad una persona, un luogo, un oggetto, etc. Sono disponibili due elementi per codificare queste espressioni:

- `<rs>` contiene un nome o un'espressione referenziale generica. Tra gli attributi:
 - *type*: indica il tipo di oggetto a cui si riferisce l'espressione. Esempi di valore possono essere *PERSONA*, *OGGETTO*, *LUOGO*, etc.
- `<name>` contiene un nome proprio o un'espressione sostantivale. Tra gli attributi:
 - *type*: indica il tipo oggetto nominato.

Uso dell'attributo *type*



L'attributo *type* viene usato, laddove sia possibile, per distinguere tra nomi di persone, luoghi ed organizzazioni:

```
<q>Mio caro <rs type='individuo'>sig. Bennet</rs>,
  </q> gli ha detto un giorno la sua signora, <q>
hai sentito che alla fine hanno affittato <rs
type='luogo'> Netherfield Park</rs>?</q>
&Egrave; uno dei principi del <rs
type='organizzazione'> Circomlocution
Office</rs>, mai, per, nessun motivo, dare una
risposta diretta,disse il <rs type='individuo'>
sig. Barnacle</rs>, <q>se possibile.</q>
```

Rapporto fra <rs> e <name>



Come mostra il seguente esempio, l'elemento <rs> può essere usato per ogni riferimento a persone, luoghi, etc., non necessariamente nella forma di un nome proprio o di un'espressione sostantivale.

```
<q> Mio caro <rs type='individuo'>sig. Bennet</rs>,
</q> gli ha detto <rs type='individuo'>la sua
signora</rs> un giorno...
```

L'elemento <name>, al contrario, viene fornito per il caso speciale di espressioni referenziali che consistono esclusivamente di nomi propri; questo può essere usato allo stesso modo dell'elemento <rs>, o inserito all'interno di questo se un'espressione referenziale contiene un insieme di nomi propri e comuni.

Uso degli attributi per i nomi



Il semplice atto di marcare qualcosa come un nome non basta - generalmente - per garantire il trattamento automatico dei nomi di persona nelle forme canoniche che di solito sono richieste per scopi di riferimento. Il nome che appare in un testo può essere scritto in maniera errata, parziale o imprecisa. Inoltre, i prefissi di nome come *di* o *de la*, possono o meno essere come parti della forma di riferimento di un nome, a seconda della lingua e dal paese di origine.

I seguenti attributi sono pertanto disponibili per questi elementi e simili, al fine di superare tali difficoltà:

- *key*: fornisce un identificatore alternativo per l'oggetto che è nominato, ad esempio una chiave per un record di database.
- *reg*: dà una forma regolarizzata o normalizzata del nome usato.

Attributo *key* e *reg*



L'attributo *key* può essere utilizzato come un mezzo per riunire tutti i riferimenti allo stesso individuo o luogo sparsi nel documento:

```
<q>Mio caro <rs type='individuo' key='BENS1'>sig. Bennet</rs>, </q> ha detto <rs type='individuo' key='BENS2'>la sua signora</rs> un giorno, <q>hai sentito che alla fine hanno affittato <rs type=luogo key='NETP1'>Netherfield Park</rs>? </q>
```

Questa utilizzazione dovrebbe essere distinta dal caso dell'attributo *reg* (regolarizzazione), che permette di segnalare la forma standard di un'espressione referenziale, come si vede nel seguente esempio:

```
<name type='individuo' key='WALDLM1' reg='de la Mare, Walter'>Walter de la Mare</name> è nato a <name key='Ch1' type='luogo'>Charlton</name>, nel <name key='KT1' type='luogo=contea'>Kent</name>, nel 1873.
```

È possibile codificare più dettagliatamente le componenti dei nomi propri, usando l'insieme di elementi supplementari per i nomi e le date.



2. Date e orari

I marcatori per la codifica dettagliata di orari e date sono i seguenti elementi:

- `<date>` contiene una data in qualche formato. Tra gli attributi:
 - *calendar*: indica il sistema o calendario al quale appartiene la data
 - *value*: dà il valore della data in qualche forma standard, ad esempio gg-mm-aaaa (giorno-mese-anno)
- `<time>` contiene un'espressione che definisce un orario in qualche formato. Tra gli attributi:
 - *value*: fornisce il valore dell'orario in una forma standard.

L'attributo *value* specifica una forma normalizzata per la data o l'orario, usando un formato riconosciuto come quello prescritto dalle norme ISO 8601. Date o orari parziali (per es. '1990', 'Settembre 1990') possono essere espressi generalmente omettendo una parte del valore fornito; invece, date o orari imprecisi (per es. 'agli inizi di Agosto', 'un po' di tempo dopo le dieci e prima di mezzogiorno') possono essere espressi come intervalli di date o orari. Se si conosce con esattezza uno dei due estremi dell'intervallo di tempo o di data (per es., 'qualche tempo prima del 1230', 'alcuni giorni dopo Carnevale') si può utilizzare l'attributo *exact* per specificarlo.



Esempio

- `<date value='21-02-1980'>21 Feb 1980</date>`
- `<date value='1990'>1990</date>`
- `<date value='09-1990'>Settembre 1990</date>`
- Dato nel `<date value='12-06-1977'>dodicesimo giorno di giugno nell'anno di Nostro Signore millenovecentosettantasette della repubblica e ottantaseiesimo dell'Università</date>`
- `<l>specialmente quando fa nove sotto zero<l> e <time value='15:00'> alle tre del pomeriggio</time>`

3. Numeri



I numeri possono essere scritti in lettere o in cifre (ventuno, XXI, 21) e la loro rappresentazione dipende dalla lingua (per esempio l'inglese *5th* diviene 5, in greco; l'inglese *123,456.78* equivale all'italiano *123.456,78*). Nelle applicazioni di trattamento del linguaggio naturale o di traduzione automatica, spesso è utile distinguere queste dalle altre parti, più 'lessicali', di un testo. In altre applicazioni, è invece importante registrare i valori numerici in una notazione standard. L'elemento `<num>` offre questa possibilità:

- `<num>` contiene un numero scritto in qualsiasi forma. Tra gli attributi:
 - *type* indica il tipo di valore numerico. I valori suggeriti sono: *FRAZIONE*, *ORDINALE* (per i numeri ordinali), *PERCENTUALE*, *CARDINALE* (per i numeri cardinali).
 - *value* specifica il valore di un numero in una forma standard specifica per un'applicazione.

Per esempio:

```
<num value='33'>xxxiii</num>
<num type=cardinale value='21'>ventuno</num>
<num type=percentuale value='10'>dieci per cento</num>
<num type=ordinale value='5'>V</num>
```

4. Abbreviazioni e loro espansioni



Come nomi, date e numeri, le abbreviazioni possono essere trascritte nella forma in cui si trovano nel testo o per esteso; esse possono essere codificate mediante i seguenti elementi:

- `<abbr>` contiene abbreviazioni di ogni tipo. Tra gli attributi:
 - *expan* fornisce l'espansione dell'abbreviazione.
 - *type* permette al codificatore di classificare l'abbreviazione secondo un'adeguata tipologia. Valori esemplificativi sono: *CONTRAZIONE*, *BREVIGRAFO*, *SOSPENSIONE*, *INTESTAZIONE*, *ACRONIMO*. L'attributo *type* può anche contenere dei valori come *TITOLO* (per i titoli di indirizzo), *GEOGRAFICO*, *ORGANIZZAZIONE*, etc, che descrivono la natura dell'oggetto a cui ci si riferisce.

L'elemento `<abbr>` è utile per distinguere elementi semi-lessicali come acronimi o termini gergali:

```
...se dopo aver eseguito un <abbr>ECG</abbr> abbiamo
un'onda Q patologica, una sopraelevazione del
segmento S-T, e un'inversione della T ci troviamo di
fronte ad un <abbr>IMA</abbr>
```

Uso dell'attributo *type*



L'attributo *type* può essere usato per distinguere i tipi di abbreviazioni a seconda delle loro funzioni, e l'attributo *expan* può essere usato per fornire un'espansione:

Il `<name>`

```
<abbr type='titolo'  expan='Dottore'>Dott.</abbr>
```

```
<abbr type='iniziale'  expan='Mario'>M.</abbr>
```

Rossi

del `<abbr expan='Centro Elaborazione Dati' type='acronimo'>CED</abbr>`

Questo elemento è particolarmente utile nel caso di trascrizioni di manoscritti, nei quali le abbreviazioni sono molto frequenti.

5. Indirizzi



L'elemento `<address>` è usato per codificare qualsiasi tipo di indirizzo postale. Contiene uno o più elementi `<addrLine>`, uno per ogni riga dell'indirizzo.

- `<address>` contiene un indirizzo postale o di altro tipo, per esempio di un individuo, di un'organizzazione, di un editore.
- `<addrLine>` contiene una riga di un indirizzo postale o di altro indirizzo.

Un semplice esempio:

```
<address>
<addrLine>Istituto Eugenio Montale</addrLine>
<addrLine>Via M. Laseppia, 35</addrLine>
<addrLine>00169 Roma</addrLine>
<addrLine>Italia</addrLine>
</address>
```

Le singole parti di un indirizzo possono essere ulteriormente distinte utilizzando l'elemento `<name>` discusso sopra nella sezione *Nomi ed espressioni referenziali*.

```
<address>
<addrLine>Istituto Eugenio Montale</addrLine>
<addrLine>Via M. Laseppia, 35</addrLine>
<addrLine>00169 <name type=cit>Roma</name></addrLine>
<addrLine><name type=sta>Italia</name></addrLine>
</address>
```

Liste



L'elemento `<list>` è usato per marcare qualsiasi tipo di lista. Una lista è una sequenza di voci testuali, che possono essere ordinate, non ordinate, o presentate sotto forma di glossario. Ogni elemento può essere preceduto da un'etichetta (in un glossario, questa etichetta è il termine che si sta definendo):

- `<list>` contiene qualsiasi sequenza di elementi organizzati come una lista. Tra gli attributi:
 - *type* descrive la forma della lista. I valori suggeriti sono: *ORDINATA*, *PUNTATA* (rispettivamente per liste con elementi numerati o in ordine alfabetico e liste con elenchi puntati), *GLOSSARIO* (per le liste costituite da una serie di termini tecnici, ognuno marcato con un elemento `<label>` e accompagnato da una glossa o da una definizione marcata come `<item>`), e *simple* (per le liste con elementi non marcati da numeri o punti).
- `<item>` contiene un componente di una lista.
- `<label>` contiene l'etichetta associata ad un elemento in una lista; nei glossari, marca il termine che si sta definendo.

Esempi



I singoli elementi di una lista sono codificati con l'elemento `<item>`. Il primo `<item>` può eventualmente essere preceduto da un elemento `<head>`, che fornisce un titolo alla lista. La numerazione di una lista può essere omessa (se ricostruibile), indicata usando l'attributo *n* per ogni elemento, o (più raramente) codificata come contenuto usando l'elemento `<label>`. I seguenti esempi sono dunque equivalenti:

```
<list>
<head>Una piccola lista</head>
<item>primo elemento in lista</item>
<item>secondo elemento in lista</item>
<item>terzo elemento in lista</item>
</list>
```

```
<list>
<head>Una piccola lista</head>
<item n=1>primo elemento in lista</item>
<item n=2>secondo elemento in lista</item>
<item n=3>terzo elemento in lista</item>
</list>
```

```
<list>
<head>Una piccola lista</head>
<label>1</label><item>primo elemento in lista</item>
<label>2</label><item>secondo elemento in lista</item>
<label>3</label><item>terzo elemento in lista</item>
</list>
```

Esempio di glossario



Una semplice tavola a due colonne può essere trattata come una lista di tipo glossario, marcata `<list type=gloss>`. In questo caso ogni elemento comprende un termine e una glossa, codificati rispettivamente con `<label>` e `<item>`. Questi corrispondono agli elementi `<term>` e `<gloss>`, che possono comparire ovunque in un testo in prosa.

```
<list type=gloss>
<head>Vocabolario</head>
<label lang='it'>ora, adesso</label><item>now</item>
<label lang='it'>rumorosamente</label><item>loudly</item>
<label lang='it'>fiori</label><item>blooms</item>
<label lang='it'>prato</label><item>meadow</item>
<label lang='it'>legno</label><item>wood</item>
<label lang='it'>pecora</label><item>ewe</item>
<label lang='it'>bassi</label><item>lows</item>
<label lang='it'>saltellare</label><item>bounds, frisks</item>
<label lang='it'>grazioso</label><item lang='fr'>joli</item>
<label lang='it'>allegrementemente</label><item>merrily</item>
<label lang='it'>smettere</label><item>cease</item>
<label lang='it'>mai</label><item>never</item>
</list>
```

Liste annidate



Naturalmente, liste di qualsiasi tipo possono essere inserite all'interno degli elementi di una lista, fino al livello di profondità necessario. Qui, per esempio, un glossario contiene due elementi, ognuno dei quali è esso stesso una semplice lista:

```
<list type='glossario'><label>SFORTUNA</label>
<item><list type='semplice'>
<item>Mi sono fratturato un braccio cadendo con la bicicletta
in un fosso.</item>
<item>Sono rimasto solo, a terra, col braccio dolente per
circa tre ore.</item>
<item>Mi sono completamente bagnato perché oltre a tutto
è cominciato a piovere.</item>
</list> <!-- fine della prima lista nidificata --></item>
<label>FORTUNA</label>
<item><list type='semplice'>
<item>Ma è passato un uomo che ha chiamato dei soccorsi.</item>
<item>Ma la frattura del braccio non era scomposta.</item>
<item>Ma cadendo da quell'altezza avrei potuto anche morire.</item>
</list><!-- fine della seconda lista nidificata --></item>
</list><!-- fine della lista glossario -->
```

Liste interne al testo



Un elenco non deve necessariamente essere presentato sotto forma di una lista. Per esempio:

Ma non v'è dubbio che la nostra preparazione sarà migliore tra qualche anno e precisamente `<list><item n='a'>`quando avremo rinnovato tutte le nostre artiglierie; `<item n='b'>`quando avremo in squadra le 8 navi da battaglia; `<item n='c'>` quando l'impero sarà del tutto pacificato, potrà bastare a se stesso e darci l'armata nera; `<item n='d'>`quando avremo realizzato almeno il 50% dei nostri piani autarchici; `<item n='e'>`a esposizione del 1942 chiusa, esposizione che deve rinforzare le nostre riserve; `<item n='f'>`quando avremo fatto rimpatriare il maggior numero possibile di italiani dalla Francia.`</list>`

(il passo riportato è ripreso da una relazione presentata da Mussolini al Gran Consiglio il 4/2/1939, ed è citato da G. Candeloro in *Storia dell'Italia moderna*, Vol. IX.)

Le liste di elementi bibliografici dovrebbero essere codificate usando l'elemento `<listBib>` descritto nella prossima sezione.

Citazioni bibliografiche



Spesso è utile marcare le citazioni bibliografiche dove esse appaiono nei testi trascritti a scopo scientifico, se non altro affinché esse siano formattate correttamente quando il testo sarà stampato. L'elemento `<bibl>` è fornito a tale fine:

`<bibl>`

contiene una citazione bibliografica non strutturata, i cui componenti possono, o meno, essere esplicitamente marcati.

Qualora i componenti di un riferimento bibliografico debbano essere distinti, altri elementi possono essere usati secondo le esigenze.

È di solito utile marcare almeno quelle parti (come titoli di articoli, libri e giornali) che necessitano di formattazioni speciali. Gli altri elementi sono forniti per casi in cui sussista un particolare interesse per tali dettagli.

Altri marcatori per le citazioni



- `<author>` in un riferimento bibliografico contiene il nome dell'autore (o degli autori), personale o collettivo, di un'opera; costituisce la dichiarazione di responsabilità principale di ciascuna unità bibliografica.
- `<biblScope>` definisce l'estensione di un riferimento bibliografico, per esempio mediante una lista di numeri di pagina, o il titolo di una parte di un'opera più ampia.
- `<date>` contiene una data in qualunque formato.
- `<editor>` dichiarazione di responsabilità secondaria per un'unità bibliografica, per esempio il nome di un individuo, istituzione od organizzazione che figura come editore, compilatore, traduttore, etc. Tra gli attributi:
 - *role*: specifica il tipo di responsabilità intellettuale. I valori esemplificativi sono: *TRADUTTORE*, *COMPILATORE*, *ILLUSTRATORE*; per default il valore è *EDITOR (CURATORE)*.
- `<imprint>` raggruppa informazioni che riguardano la pubblicazione o la distribuzione di un'unità bibliografica.
- `<publisher>` fornisce il nome dell'organizzazione responsabile della pubblicazione, o distribuzione di un'unità bibliografica.
- `<pubPlace>` contiene il nome del luogo in cui è stata pubblicata un'unità bibliografica.
- `<series>` contiene indicazioni sulla collana in cui un libro o un'altra unità bibliografica sono stati pubblicati.
- `<title>` contiene il titolo di un'opera, sia essa un articolo, un libro, un giornale, o una collana, che includa qualunque titolo alternativo o sottotitolo. Tra gli attributi:
 - *type*: categorizza il titolo, per esempio come principale, subordinato, etc.
 - *level*: indica il livello bibliografico o la classe del titolo. I valori permessi sono descritti nella sezione Cambiamenti degli stili di carattere.

Esempio



Per esempio la seguente nota editoriale potrebbe esser trascritta come segue:

Vedi anche il frammento del 1345 pubblicato dal
Livi, *Dante* 50-52

```
Vedi anche il frammento del 1345 pubblicato dal  
<bibl><author>Livi</author> <title>Dante</title>  
<biblScope>50-52</biblScope></bibl>.
```

Per le liste di citazioni bibliografiche andrebbe usato l'elemento `<listBibl>`, che può contenere una serie di elementi `<bibl>`. Per un esempio rinviamo alla lista della sezione *Riferimenti bibliografici*.

Tavole e tabelle



Le tabelle rappresentano una sfida considerevole per qualsiasi sistema di trattamento del testo; tuttavia, semplici tabelle, come minimo, compaiono in così tanti testi, che risulta necessario inserire un sistema per la loro rappresentazione anche nell'insieme di elementi TEI semplificato qui esposto. I seguenti elementi sono provvisti a tale scopo:

- `<table>` contiene del testo presentato in forma di tabella, in righe e colonne. Tra gli attributi:
 - `rows`: indica il numero di righe della tabella.
 - `cols`: indica il numero di colonne di ogni riga della tabella.
- `<row>` contiene una riga della tabella. Tra gli attributi:
 - `role`: indica il tipo d'informazione contenuta nelle caselle di una riga. I valori consigliati includono `LABEL`, per le etichette o per informazioni descrittive e `DATA`, per gli effettivi valori dei dati.
- `<cell>` contiene una casella di una tabella. Tra gli attributi:
 - `role`: indica il tipo di informazione contenuta nella casella. I valori consigliati includono `LABEL`, per le etichette o per informazioni descrittive e `DATA`, per gli effettivi valori dei dati.
 - `cols`: indica il numero di colonne occupate da questa casella.
 - `rows`: indica il numero di righe occupate da questa casella.

Esempio



Per esempio in questa tavola ripresa dal manuale di storia *L'età contemporanea* di A. Giardina, G. Sabbatucci, V. Vidotto viene mostrata la crescita della popolazione, in migliaia di unità, di alcune delle principali città europee nel corso del XIX secolo:

```
<table rows='5' cols='5'>
<row='data'>
<cell role='label'>anno</cell>
<cell>1800</cell><cell>1850</cell><cell>1880</cell><cell>1910</cell></row>
<cell role='label'>Amburgo</cell>
<cell>130</cell><cell>132</cell><cell>290</cell><cell>932</cell></row>
<cell role='label'>Berlino</cell>
<cell><172></cell><cell>419</cell><cell>1122</cell><cell>2071</cell></row>
<cell role='label'><cell>Budapest</cell>
<cell>54</cell><cell>178</cell><cell>371</cell><cell>880</cell></row>
<cell role='label'><cell>Costantinopoli</cell>
<cell>600</cell><cell>&mdash;</cell><cell>&mdash;</cell><cell>1200</cell>
</row>
</table>
```

Immagini e grafica



Non tutti i componenti di un documento sono necessariamente testuali. Anche il testo più semplice spesso contiene diagrammi o illustrazioni, per non parlare dei documenti in cui immagine e testo sono inestricabilmente connessi, o di risorse elettroniche in cui i due sono complementari.

Il codificatore può semplicemente registrare la presenza di un elemento grafico all'interno di un testo, eventualmente con una breve descrizione del suo contenuto, usando gli elementi descritti in questa sezione. Gli stessi elementi possono inoltre essere usati per inserire versioni digitalizzate di elementi grafici all'interno di un documento elettronico.

`<figure>` indica il luogo in cui un elemento grafico deve essere inserito in un documento. Tra gli attributi:

`entity` il nome di un'entità di sistema predefinita che contiene una versione digitalizzata dell'immagine da inserire.

`<figDesc>` contiene una descrizione testuale dell'aspetto o del contenuto di un elemento grafico, utilizzabile quando si documenta un'immagine senza visualizzarla.

Esempi



Tutta l'informazione testuale che accompagna l'elemento grafico, come un titolo e/o una didascalia, deve essere inclusa all'interno dell'elemento `<figure>`, in un elemento `<head>` ed in uno o più elementi `<p>`, così come qualsiasi altro testo che appaia al suo interno. Si consiglia fortemente di fornire una descrizione testuale dell'immagine, nell'ambito di un elemento `<figDesc>`. Tali informazioni potranno essere usate da applicazioni che non sono in grado di visualizzare immagini e renderanno accessibile il documento ai lettori con problemi alla vista (la descrizione non è di solito considerata parte del documento).

La più semplice applicazione di questi elementi consiste nel marcare la posizione di un'immagine, come nell'esempio:

```
<pb n='413' /> <figure></figure> <pb n='414' />
```

Più spesso un'immagine ha un titolo d'identificazione che deve essere codificato usando l'elemento `<head>`. Inoltre, è spesso conveniente includere una breve descrizione dell'immagine come nel seguente esempio:

```
<figure>
<head>Tobiolo e l'angelo</head>
<figDesc>raffigura il passo biblico (Tob. VI, 2-5), in cui si narra
che Tobiolo e l'angelo decisero di passare la notte presso il fiume
Tigri.</figDesc>
</figure>
```

Inserimento di immagini e grafici



Quando è disponibile una versione digitalizzata di un elemento grafico, è chiaramente preferibile includerla nel punto appropriato all'interno del testo. Gli elementi grafici, come le fotografie, sono normalmente memorizzate entità (file) distinte da quelle che contengono il testo di un documento, e si servono di una differente notazione (formato di memorizzazione). La DTD della *TEI Lite* supporta immagini codificate nei formati standard CGM, TIFF, e JPEG con i nomi di notazione XML *cgm*, *tiff*, e *jpeg* rispettivamente. Altre notazioni possono essere comunque usate, a condizione che un'appropriata dichiarazione NOTATION venga aggiunta alla DTD (per maggiori informazioni cfr. *TEI P4*).

Qualunque formato venga usato per codificare l'immagine, essa può essere inserita all'interno del documento nella stessa maniera. Il primo passo consiste nella dichiarazione di una particolare entità XML, che specifica un nome per l'entità, un identificatore esterno (come il nome del file) e la notazione usata. Per esempio, ipotizzando che l'immagine digitalizzata di Tobio fosse stata inserita in formato TIFF nel file *tobi.tif*, sarebbe necessaria una dichiarazione d'entità come la seguente:

```
<!ENTITY tobiFig SYSTEM 'tobi.tif' NDATA tiff>
```

Ancora sull'inserimento delle immagini



Qualunque dichiarazione deve essere elaborata prima del documento XML stesso; con la DTD della *TEI Lite* ciò è possibile includendole in un file chiamato *figures.ent* o qualunque altro file associato all'identificatore pubblico `-//TEI U5-1995//DTD TEI Lite 1.0 Extensions//EN`.

Una volta che la dichiarazione sopra menzionata è attiva, tutto ciò che è necessario per inserire l'immagine digitalizzata nel punto appropriato del documento è fornire un valore all'attributo *entity* dell'elemento `<figure>`.

```
<figure entity='tobiFig'>
<head>Tobiolo e l'angelo</head>
<figDesc>raffigura il passo biblico (Tob. VI, 2-5), in
cui si narra che Tobiolo e l'angelo decisero di
passare la notte presso il fiume Tigri.</figDesc>
</figure>
```

Interpretazione ed analisi



Si dice spesso che la codifica testuale è una forma d'interpretazione o di analisi. Sebbene sia certamente difficile, se non impossibile, distinguere con certezza tra informazioni 'oggettive' e 'soggettive' in un modo universalmente valido, non vi è dubbio sul fatto che i giudizi relativi alle seconde siano, di norma, più controversi rispetto a quelli relativi alle prime. Molti studiosi, pertanto, preferiscono presentare tali interpretazioni soggettive solo se è possibile indicare esplicitamente al lettore che si tratta di caratteristiche testuali da considerare con maggiore cautela rispetto al resto della codifica. Questa sezione descrive alcuni degli elementi provvisti dallo schema TEI per far fronte a questa esigenza.

1. Frasi ortografiche
2. Elementi generici di interpretazione

1. Frasi ortografiche



Solitamente, l'interpretazione spazia attraverso un intero testo, senza tenere in particolare conto le altre unità strutturali. Un preliminare utile per inserire intensivamente interpretazioni consiste pertanto nel segmentare il testo in unità discrete ed identificabili, ognuna delle quali potrà poi ricevere un'etichetta da usare come una sorta di 'riferimento canonico'. Al fine di facilitare questo tipo di utilizzazione, tali unità non possono sovrapporsi, né includersi l'una dentro l'altra. Esse possono essere rappresentate adeguatamente mediante il seguente elemento:

<s> identifica una *s-unità* all'interno di un documento, al fine di costituire uno schema di riferimento canonico che si riferisca all'intero testo. Tra gli attributi:

- *type*: categorizza l'unità (per es. *DICHIARATIVA*, *INTERROGATIVA*, etc.)

Ancora sulle frasi ortografiche



L'elemento `<s>` è usato comunemente (soprattutto nelle applicazioni linguistiche) per marcare le frasi ortografiche, ovvero unità definite da caratteristiche ortografiche come l'interpunzione. Per esempio il passaggio da *Il lanciatore di Giavellotto* discusso prima può essere diviso in *s-unità* come segue:

```
<pb n='163' />
<div1 type='capitolo' n='16'>
<p><s n='001'>&mdash; <q>Sono contenta che tu sia bravo,</q> &mdash;</s>
<s n='002'>infine poté dire quietamente la madre.</s>
<p><s n='003'>Dam&iacute;n sorrise.</s>
<p><s n='004'>&mdash; <q>Bravo tanto da impressionare i professori.</s>
<s n='005'>Tanto bravo da diventare un artista.</s>
<s n='006'>Te lo meriti perché sei buono;</s>
<s n='007'>e se lo merita anche tuo nonno.</s>
<s n='008'>Anche lui è un artista; anche se è rimasto a fare cocci...
un vero artista.</s>
<s n='009'>Chissà anche lui come sarà contento.</s>
<s n='010'>Il segno D.P. continua;</s>
<s n='011'>continua anche nell'arte, come ha detto il professore.</s></q>
```

Quando le *s-unità* sono codificate come mostrato sopra è consigliabile marcare l'intero testo sino alla fine, cosicché ogni parola nel testo analizzato sarà inclusa in un solo elemento `<s>`, il cui identificatore può successivamente essere usato come sistema di riferimento. Se gli identificatori impiegati in un documento sono unici, allora può essere usato l'attributo *id*, preferibile all'attributo *n* usato nell'esempio precedente.

2. Elementi generici di interpretazione



Un elemento di segmentazione più generico, il `<seg>`, è già stato introdotto: serve per identificare le destinazioni di riferimenti incrociati e collegamenti ipertestuali non altrimenti marcate (vedere la sezione *Riferimenti incrociati e collegamenti*); esso identifica qualsiasi espressione di livello frasale alla quale il codificatore può assegnare un *type* specificato dall'utente, ed un identificatore unico; può in questo modo essere usato per codificare caratteristiche testuali che non sono previste nelle *Norme TEI*.

Per esempio, le *Norme* non prevedono nessun elemento `<apostrophe>` per marcare parti di un testo letterario in cui il narratore si rivolge direttamente al lettore (o all'ascoltatore). Un modo di affrontare il problema è considerare queste come istanze dell'elemento `<q>`, distinte dalle altre tramite un valore appropriato per l'attributo *who*. Una soluzione più semplice, e certamente più generale, invece, potrebbe essere l'uso dell'elemento `<seg>` come segue:

```
<p> <seg type='apostrophe'>E insomma, lo volete fare
anche voi, sì o no, questo esperimento con me, una
buona volta?</seg>
```

(il passo è tratto da *Uno, nessuno e centomila* di L. Pirandello)

Elemento <seg>



L'attributo *type* dell'elemento <seg> può assumere qualsiasi valore, e può così essere usato per registrare fenomeni espressivi di qualunque genere; è bene registrare i valori usati ed il loro significato nella testata.

Un elemento <seg> di un dato tipo (diversamente dall'elemento <s> che gli somiglia superficialmente) può essere annidato all'interno di un elemento <seg> dello stesso tipo o di tipo differente.

Questo consente di rappresentare strutture abbastanza complesse; alcuni esempi sono stati forniti nella sezione *Attributi di collegamento*.

Tuttavia, poiché esso deve rispettare i vincoli XML secondo cui gli elementi devono essere propriamente annidati e non possono incrociarsi tra loro, questo elemento non può rispondere all'esigenza di associare un'interpretazione a segmenti arbitrari di un testo, che potrebbero ignorare completamente la gerarchia del documento. Esso inoltre richiede che l'interpretazione stessa venga rappresentata da un valore singolo, codificato nell'attributo *type*.

Elemento <interp>



Nessuna di tali restrizioni si applica all'elemento <interp>, che offre un sistema molto potente per codificare interpretazioni assai complesse in maniera relativamente semplice.

- <interp> fornisce un'annotazione interpretativa che può essere collegata a parti di testo. Tra gli attributi:
 - *value*: identifica lo specifico fenomeno che è stato annotato.
 - *resp*: indica chi è responsabile dell'interpretazione.
 - *type*: indica il tipo di fenomeno notato nel brano. Valori esemplificativi sono *IMMAGINE*, *PERSONAGGIO*, *TEMA*, *ALLUSIONE*, o il nome di un particolare genere di discorso i cui esempi sono identificati.
 - *inst*: punta alle istanze dell'analisi o dell'interpretazione rappresentate dall'elemento corrente.
- <interpGrp> raggruppa un insieme di elementi <interp>.

Uso degli elementi di interpretazione



Questi elementi permettono al codificatore di specificare sia un classe di interpretazione, sia il particolare esempio della classe implicata dall'interpretazione. Così, mentre con `<seg>` si può dire solamente che qualcosa è un'apostrofe, con `<interp>` si può dire che si tratta di un esemplare (apostrofe) di una classe più ampia (quella delle figure retoriche).

Inoltre, `<interp>` è un elemento vuoto che deve essere collegato al passaggio a cui si applica o tramite l'attributo `ana` discusso in precedenza nella sezione *Attributi di collegamento*, oppure tramite il suo stesso attributo `inst`. Questo significa che si può rappresentare qualsiasi tipo d'analisi, senza bisogno di rispettare la gerarchia del documento XML, ed inoltre facilita il raggruppamento di analisi di un dato tipo. L'elemento specifico `<interpGrp>` è fornita per quest'ultimo scopo.

Prendiamo per esempio, questo brano tratto da *La lupa* di Giovanni Verga:

```
<p>Al villaggio la chiamavano <emph rend='it'>la Lupa</emph> perché non era sazia giammai &mdash;di nulla. Le donne si facevano la croce quando la vedevano passare, sola come una cagnaccia, con quell'andare randagio e sospettoso della lupa affamata; ella si spolpava i loro figlioli e i loro mariti in un batter d'occhio, con le sue labbra rosse, e se li tirava dietro alla gonnella solamente a guardarli con quegli occhi da satanasso, fossero stati davanti all'altare di Santa Agrippina. Per fortuna <emph rend='it'>la Lupa</emph> non veniva mai in chiesa né a Pasqua, né a Natale, né per ascoltare messa, né per confessarsi. &mdash; Padre Angiolino di Santa Maria di Gesù, un vero servo di Dio, aveva persa l'anima per lei.</p>
```

Interpretazione: figure retoriche



Supponiamo che si vogliano marcare diversi aspetti del testo, come temi o soggetti, figure retoriche e i luoghi di singole scene della narrazione. Le differenti parti del nostro brano, per esempio, potrebbero essere associate alle figure retoriche dell'isotopia, dell'iperbole e della metafora; con riferimenti per soggetto alla protagonista, alle donne, all'altare e con scene site nel villaggio ed in luoghi non specificati.

Queste interpretazioni possono essere messe ovunque all'interno dell'elemento `<text>`; è comunque buona pratica inserirle tutte nello stesso posto (per esempio in una sezione separata dell'avantesto o degli annessi), come nel seguente esempio:

```
<back>
<divl type='Interpretazione'>
<interp id='fig-met' resp='MG, AL' type='figura della protagonista'
value='metafora' />
<interp id='fig-ipe' resp='MG, AL' type='figura della protagonista'
value='iperbole' />
<interp id='fig-iso' resp='MG, AL' type='figura della protagonista'
value='isotopia' />
<!-- ... -->
<interp id='scen-villaggio' resp='MG, AL' type='scenario' value='villaggio' />
<!-- ... -->
<interp id='ref-prot' resp='MG, AL' type='referenza' value='protagonista' />
<interp id='ref-don' resp='MG, AL' type='referenza' value='donne' />
<!-- ... -->
</p>
</divl>
```

Uso dell'elemento <interGrp>



L'evidente ridondanza di questa codifica può essere considerevolmente ridotta usando l'elemento <interpGrp> per raggruppare tutti quegli elementi <interp> che condividono valori di attributi comuni, come segue:

```
<back>
<div1 type='Interpretazione'>
  <interpGrp type='figura della protagonista' resp='MG, AL'>
    <interp id='fig-met' value='metafora' />
    <interp id='fig-ipe' value='iperbole' />
    <interp id='fig-iso' value='isotopia' />
    <!-- ... -->
  </interpGrp>
  <interpGrp type='scenario' resp='MG, AL'>
    <interp id='scen-villaggio' value='villaggio' />
    <interp id='scen-non spec' value='non-specificato' />
    <!-- ... -->
  </interpGrp>
  <interpGrp type='referenza' resp='MG, AL'>
    <interp id='ref-prot' value='protagonista' />
    <interp id='ref-don' value='donne' />
    <!-- ... -->
  </interpGrp>
</div1>
```

Collegamento interpretazione/testo



Una volta che queste interpretazioni sono state definite, possono essere collegate con le parti del testo a cui si riferiscono, in uno dei modi o in entrambi. L'attributo *ana* può essere usato per qualsivoglia elemento appropriato:

```
<div1 type='paragrafo' n='2'>
  <p id='P2' ana='scen-villaggio'>
    <s id='P2.1' ana='fig-met'>Al villaggio la chiamavano <emph
      rend='italic'>la Lupa</emph> perché non era sazia giammai
      &mdash; di nulla.</s> ...
```

Altrimenti, gli elementi <interp> possono indicare tutte le parti del testo a cui si riferiscono, usando il loro attributo *inst*:

```
<interp id='fig-met' type='figura della protagonista' resp='MG,
  AL' value='metafora' inst='P2.1' />
<!-- ... -->
<interp id='scen=villaggio' type='scenario' value='villaggio'
  inst='P2' resp='MG, AL' />
<interp id='scen=non-specifico' type='scenario' value='non-
  specifico' inst='P2' resp='MG, AL' />
<!-- ... -->
```

Altri usi di <interp> per l'analisi



L'elemento <interp> non è limitato a un tipo particolare di analisi. L'analisi letteraria mostrata sopra rappresenta una delle possibilità; si può egualmente usare <interp> per fornire un'analisi linguistica del discorso. Per esempio, la frase data nella sezione *Attributi di collegamento* ha un'analisi linguistica che potrebbe essere rappresentata come segue:

```
<interp id='NP1' type='pos' value='sintagma-  
nominale, singolare' />  
<interp id='VV1' type='pos'  
value='coniugazione verbo, terza pers.  
singolare tempo presente' />
```

Elementi dell'avantesto



Per numerose applicazioni, specialmente nel caso di testi antichi, le pagine preliminari di un volume, come il frontespizio, le lettere di prefazione, etc., possono fornire informazioni utilissime di tipo linguistico e sociologico. La *TEI* dà un insieme di suggerimenti per distinguere gli elementi testuali che si incontrano solitamente nell'avantesto, che vengono qui riassunti.

1. Frontespizio
2. Materiali introduttivi

1. Frontespizio



L'inizio della pagina di frontespizio dovrebbe essere codificato con l'elemento `<titlePage>`. Tutto il testo contenuto nella pagina andrebbe trascritto e marcato appropriatamente con uno tra gli elementi qui elencati:

- `<titlePage>` contiene la pagina di frontespizio di un testo, che appare all'interno dell'avantesto o tra gli annessi.
- `<docTitle>` contiene il titolo di un documento, tutti i suoi elementi costitutivi, come appaiono sul frontespizio; deve essere diviso in elementi `<titlePar>`.
- `<titlePart>` contiene una suddivisione del titolo di un'opera, come appare sul frontespizio; va inoltre usato per frammenti senza collocazione fissa del frontespizio che non fanno parte del titolo, attribuzioni di paternità, etc. Tra gli attributi:
 - *type* precisa il ruolo di questa suddivisione del titolo. I valori suggeriti includono: *TITOLO PRINCIPALE*, *SOTTOTITOLO*, *DESCRIZIONE* (una parafrasi descrittiva dell'opera inclusa nel titolo), *ALTERNATIVO* (titolo alternativo).
- `<byline>` contiene la dichiarazione di responsabilità primaria di un'opera, che si trova sul suo frontespizio oppure all'inizio o alla fine dell'opera.
- `<docAuthor>` contiene il nome dell'autore del documento, così come è dato sul frontespizio (spesso, ma non sempre, inserito in un elemento `<byline>`).
- `<docDate>` contiene la data di edizione del documento, come indicata (abitualmente) sul frontespizio.
- `<docEdition>` contiene la dichiarazione dell'edizione, così come è presentata sul frontespizio di un documento.
- `<docImprint>` contiene la dichiarazione delle note tipografiche (luogo e data di pubblicazione, nome dell'editore), così come appare (abitualmente) alla base del frontespizio.
- `<epigraph>` contiene una citazione, anonima o attribuita, che appare all'inizio di una sezione o capitolo, o sul frontespizio.

Altri elementi del frontespizio



Le differenziazioni dei caratteri tipografici dovrebbero essere marcate con l'attributo *rend* quando è necessario, come precedentemente descritto. Una descrizione molto dettagliata della spaziatura e delle dimensioni dei caratteri usati nei titoli ornamentali non è stata ancora prevista dalle *Norme TEI*. I cambiamenti di lingua devono essere marcati con l'uso appropriato dell'attributo *lang* o dall'elemento `<foreign>`, a seconda dei casi. I nomi, qualora apparissero, devono essere marcati usando l'elemento `<name>`, come altrove.

Esempio



```
<titlePage>
  <docImprint>
    <publisher>
      <name>Universale Economica Feltrinelli</name>
    </publisher>
    <pubPlace>
      <name>Milano</name>
    </pubPlace>
  </docImprint>
  <byline>
    <docAuthor>Giorgio Candeloro</docAuthor>
  </byline>
  <docTitle>
    <titlePart type='main'>STORIA DELL'ITALIA MODERNA</titlePart>
    <titlePart>nono <hi>volume</hi></titlePart>
    <titlePart type='sub'>IL FASCISMO E LE SUE GUERRE</titlePart>
  </docTitle>
  <docDate>MCMXCV</docDate>
</titlePage>
```

Altro esempio



```
<titlePage>
  <docTitle>
    <titlePart type='main'> Lives of the Queens of England, from
    the Norman Conquest; </titlePart>
    <titlePart type='sub'>with anecdotes of their courts.
    </titlePart>
  </docTitle>
  <titlePart>Now first published from Official Records and other
  authentic documents private as well as public. </titlePart>
  <docEdition>New edition, with corrections and
  additions</docEdition>
  <byline>By
    <docAuthor>Agnes Strickland</docAuthor>
  </byline>
  <epigraph>
    <q>The treasures of antiquity laid up in old historic rolls,
    I opened.</q>
    <bibl>BEAUMONT</bibl>
  </epigraph>
  <docImprint>Philadelphia: Blanchard and Lea</docImprint>
  <docDate>1860.</docDate>
</titlePage>
```

2. Materiali introduttivi



Le principali sezioni testuali all'interno dell'avantesto dovrebbero essere codificate come elementi `<div>` o `<div1>`; i seguenti valori suggeriti per l'attributo `type` possono servire per distinguere i vari tipi comuni di materiali introduttivi:

- **PREFAZIONE** un testo indirizzato al lettore, dall'autore, curatore o editore, eventualmente in forma di lettera.
- **INTRODUZIONE** un testo indirizzato al lettore, dall'autore, curatore o editore, eventualmente in forma di lettera.
- **DEDICA** un testo (spesso una lettera) indirizzato a qualcuno che non sia il lettore generico, in cui l'autore solitamente offre l'opera all'attenzione della persona in questione.
- **ABSTRACT** un riassunto in prosa del contenuto dell'opera.
- **RINGRAZIAMENTI** i ringraziamenti
- **SOMMARIO** un sommario (di solito da codificare come `<list>`)
- **FRONTESPIZIO** una decorazione del frontespizio, eventualmente corredata da testo.

Marcatori per materiali introduttivi



Come qualsiasi altra sezione del testo, quelle situate nell'avantesto possono contenere elementi di livello strutturale inferiore, o elementi non strutturali, come descritto altrove. Esse inizieranno in genere con un'intestazione o un titolo di qualche tipo, che deve essere marcato mediante l'elemento `<head>`. Le epistole conterranno i seguenti elementi supplementari:

- `<salute>` contiene una formula o un saluto di apertura in una prefazione, un'epistola dedicatoria o altre sezioni di un testo, o una formula di saluto alla fine di una lettera, prefazione, etc.
- `<signed>` contiene la formula di commiato apposta alla fine di una prefazione, di un'epistola dedicatoria, o di altre divisioni di un testo.
- `<byline>` contiene la dichiarazione di responsabilità primaria di un'opera, che si trova sul suo frontespizio oppure all'inizio o alla fine dell'opera.
- `<dateline>` contiene una breve descrizione del luogo, data, periodo, etc. di produzione di una lettera, fatto di cronaca o altre opere, messa all'inizio o alla fine di questi, a mo' di titolo o di chiusura.
- `<argument>` una lista formale o una descrizione in prosa dei temi trattati in una sezione del testo.
- `<cit>` una citazione da qualche altro documento, insieme ad un riferimento bibliografico alla sua fonte.
- `<opener>` raggruppa insieme date, firme, saluti ed espressioni simili che appaiono come un gruppo preliminare all'inizio di una sezione, specialmente se si tratta di una lettera.
- `<closer>` raggruppa insieme date, firme, saluti ed espressioni simili che appaiono come un gruppo preliminare alla fine di una sezione, specialmente se si tratta di una lettera.

Esempio



Le epistole che appaiono altrove in un testo conterranno, ovviamente, questi medesimi elementi.

Come esempio, la dedica all'inizio dell'*Epistola XIII* a Cangrande della Scala di Dante, andrebbe codificata come segue:

```
<div type='dedication'>
<opener>
<head>Al magnifico e vittorioso signore, il signore
  <name>Cangrande della Scala</name>, Vicario generale del
  sacratissimo Cesareo Principato nella città di Verona e
  nella città di Vicenza,</head>
<byline><name>Dante Alighieri</name>, fiorentino di nascita
  non di costumi, a lui devotissimo,</byline>
<salute>augura vita felice per lunghi anni e che la gloriosa
  rinomanza possa accrescersi per l'eternità.</salute>
</opener>
...
</div>
```

Elementi degli annessi. Divisioni strutturali



A causa dei cambiamenti intervenuti nella pratica editoriale, le parti finali o annessi di un volume possono contenere virtualmente tutti gli elementi elencati sopra per l'avantesto, e gli stessi elementi andrebbero usati laddove si presenti lo stesso fenomeno. Inoltre, gli annessi possono contenere i seguenti tipi di materiali, inseriti all'interno dell'elemento *<back>*. Come avviene per le divisioni strutturali del corpo del testo, essi dovrebbero essere codificati come elementi *<div>* o *<div1>*, e distinti mediante i seguenti valori consigliati dell'attributo *type*:

- **APPENDICE** un'appendice
- **GLOSSARIO** una lista di parole e definizioni, solitamente nella forma di una lista di tipo glossario.
- **NOTE** una serie di note (*<note>*)
- **BIBLIOGRAFIA** una serie di riferimenti bibliografici, solitamente nella forma dello speciale elemento lista bibliografica *<listBibl>*, le cui voci sono singoli elementi *<bibl>*.
- **INDICE** un insieme di entrate di indice, eventualmente rappresentate come una lista strutturata o una lista di tipo glossario, con un'introduzione opzionali *<head>*, ed eventualmente alcuni paragrafi di testo introduttivo o di chiusura (la *TEI P3* definisce ulteriori elementi specializzati per generare indici nella produzione di documenti).
- **COLOPHON** una descrizione sul retro del libro che riporta dove, quando e da chi è stato stampato; nei libri moderni esso spesso contiene dettagli di produzione ed identifica i caratteri tipografici usati.

Il Frontespizio elettronico



Ogni testo TEI ha una testata che offre informazioni analoghe a quelle fornite dal frontespizio di un testo a stampa. La testata è introdotta dall'elemento `<teiHeader>` ed è composta da quattro parti principali:

1. `<fileDesc>` contiene una descrizione bibliografica completa di un file digitale.
2. `<encodingDesc>` documenta le relazioni tra un testo elettronico e la fonte, o le fonti, da cui è stato tratto.
3. `<profileDesc>` contiene una descrizione dettagliata degli aspetti non bibliografici di un testo, specificamente le lingue e i dialetti usati, le circostanze in cui è stato prodotto, i partecipanti e il loro ambiente.
4. `<revisionDesc>` riassume la storia delle revisioni di un documento elettronico.

Un corpus o una collezione di testi, che condividono molte caratteristiche, possono avere una testata comune per il corpus, e testate individuali per ogni componente del corpus stesso. In questo caso l'attributo `type` indica il tipo di testata.

`<teiHeader type='corpus'>` introduce la testata per le informazioni relative al corpus.

Elementi interni alle partizioni



Alcuni elementi della testata contengono semplicemente del testo in prosa, codificato come uno o più elementi `<p>`. Altri sono raggruppati:

- * gli elementi i cui nomi finiscono in *Stmt* (per *statement*) contengono di solito un gruppo di elementi che contengono informazioni strutturate.
- * gli elementi i cui nomi finiscono in *Decl* (per *declaration*) includono informazioni sulle specifiche pratiche di codifica messe in atto nel documento.
- * gli elementi i cui nomi finiscono in *Desc* (per *description*) contengono una descrizione in testo libero.

1. <fileDesc> Descrizione del file



L'elemento <fileDesc> è obbligatorio. Contiene una descrizione bibliografica completa del file, mediante i seguenti elementi:

- a. <titleStmt> raggruppa le informazioni sul titolo di un'opera e sui responsabili del suo contenuto intellettuale.
- b. <editionStmt> raggruppa le informazioni relative ad una data edizione del testo.
- c. <extent> descrive le dimensioni approssimative del testo elettronico, memorizzato su un dato supporto, espresse in una qualsiasi unità di misura adeguata.
- d. <publicationStmt> raggruppa le informazioni relative alla pubblicazione e distribuzione di un testo, elettronico o meno.
- e. <seriesStmt> raggruppa le informazioni sulla collana, se esiste, alla quale appartiene la pubblicazione.
- f. <notesStmt> raggruppa tutte le note che forniscono informazioni supplementari su un testo oltre a quelle registrate in altre parti della descrizione bibliografica.
- g. <sourceDesc> fornisce una descrizione bibliografica del testo (o dei testi) da cui è stato tratto o generato un testo elettronico.

Elementi minimali per <fileDesc>



Una testata ha al minimo la seguente struttura:

```
<teiHeader>
  <fileDesc>
    <titleStmt> ... </titleStmt>
    <publicationStmt> ...
  </publicationStmt>
    <sourceDesc> ... <sourceDesc>
  </fileDesc>
</teiHeader>
```

1a. Dichiarazione del titolo



Un elemento *<titleStmt>* può contenere i seguenti elementi:

- *<title>* contiene il titolo di un'opera, sia essa articolo, libro, giornale o collana, tutti i titoli alternativi o i sottotitoli.
- *<author>* in un riferimento bibliografico, contiene il nome dell'autore/i, personale o collettivo, di un'opera; costituisce la dichiarazione di responsabilità primaria per ogni unità bibliografica.
- *<sponsor>* specifica il nome dell'organizzazione o istituzione che sponsorizza un progetto di ricerca.
- *<funder>* specifica il nome di un individuo, istituzione od organizzazione che ha finanziato la realizzazione del progetto o del testo.
- *<principal>* fornisce il nome del principale ricercatore responsabile della creazione di un testo elettronico.
- *<respStmt>* fornisce una dichiarazione di responsabilità relativa al responsabile del contenuto intellettuale di un testo, un'edizione, una registrazione, una collana, se gli elementi specifici per autori, curatori, etc., non sono sufficienti o non sono adatti.

Specificazioni sull'uso dell'elemento



È consigliabile che il titolo del documento elettronico sia diverso da quello del testo di origine, per esempio:

[titolo della fonte]: trascrizione linguaggio macchina

[titolo della fonte]: edizione elettronica

Una versione machine-readable di: [titolo della fonte]

Sottopartizioni di <respStmt>



L'elemento <respStmt> contiene i seguenti sotto-componenti:

- <resp> contiene un'espressione che descrive per esteso la natura della responsabilità intellettuale di una persona.
- <name> contiene un nome proprio o un'espressione sostantivale.

Esempio:

```
<titleStmt>
  <title>Il turno: edizione elettronica</title>
  <author>Luigi Pirandello</author>
  <respStmt>
    <resp>Edizione elettronica a cura del
    dott.</resp> <name>Fabio Ciotti</name>
  </respStmt>
</titleStmt>
```

1b. Dichiarazione dell'edizione



L'elemento <editionStmt> raggruppa le informazioni relative ad una data edizione di un testo (dove edizione è usato nell'accezione bibliografica comune del termine) e potrebbe includere i seguenti elementi:

- <edition> descrive le particolarità di un'edizione di un testo.
- <respStmt> fornisce una dichiarazione di responsabilità relativa al responsabile del contenuto intellettuale di un testo, un'edizione, una registrazione, una collana, se gli elementi specifici per autori, curatori, etc., non sono sufficienti o non sono adatti.

Esempio:

```
<editionStmt>
  <edition n='U2'>terza bozza sostanzialmente
  revisionata
  <date>1987</date>
</edition>
</editionStmt>
```

La determinazione esatta di cosa costituisca una nuova edizione di un testo elettronico è lasciata al giudizio di colui che effettua la codifica.

1c. Dichiarazione della dimensione



La dichiarazione `<extent>` descrive le dimensioni approssimative di un file.

Esempio:

```
<extent>4532 bytes</extent>
```

1d. Dichiarazione della pubblicazione



L'elemento `<publicationStmt>` è obbligatorio. Può contenere una semplice descrizione in prosa o gruppi dei seguenti elementi:

- `<publisher>` fornisce il nome dell'organizzazione responsabile della pubblicazione, o distribuzione di un'unità bibliografica.
- `<distributor>` fornisce il nome di una persona o di un'organizzazione responsabile della distribuzione di un documento.
- `<authority>` fornisce il nome di una persona o di un'organizzazione responsabile della messa a disposizione di un file elettronico, quando non coincidano con l'editore o il distributore.

1d. Dichiarazione della pubblicazione – altri elementi



Almeno uno di questi tre elementi deve essere presente, a meno che tutte le informazioni sulla pubblicazione non siano fornite in testo libero.

Questi elementi possono a loro volta contenere i seguenti elementi:

- `<pubPlace>` contiene il nome del luogo di pubblicazione di un documento.
- `<address>` contiene un indirizzo postale o di altro genere, per esempio di un editore, di un'organizzazione o di un singolo individuo.
- `<idno>` fornisce ogni numero, standardizzato o meno, usato per identificare un'unità bibliografica. Tra gli attributi:
 - *type* categorizza il numero, per esempio come una serie ISBN o altre serie standardizzate.
- `<availability>` fornisce informazioni sulla disponibilità di un testo, per esempio su qualsiasi restrizione che si applica alla sua utilizzazione o distribuzione, il suo statuto in merito al copyright, etc. Tra gli attributi:
 - *status* fornisce un codice che identifica la disponibilità effettiva di un testo. I valori esemplificativi sono: *RISTRETTO*, *SCONOSCIUTO*, *LIBERO*.
- `<date>` contiene una data in qualunque formato.

1d. Dichiarazione della pubblicazione – esempio



```
<publicationStmt>
  <publisher>Università Tor Vergata Roma</publisher>
  <pubPlace>Roma</pubPlace>
  <date>1996</date>
  <idno type='ISBN'>3-5-967584-4</idno>
  <availability>
    Copyright 1996, Università Tor Vergata Roma
  </availability>
</publicationStmt>
```

1e. Dichiarazione di collane e note



L'elemento `<seriesStmt>` contiene informazioni sulla collana, se esiste, a cui appartiene una pubblicazione. Può contenere `<title>`, `<idno>` o `<respStmt>`.

L'elemento `<noteStmt>`, se usato, contiene uno o più elementi `<note>` che a loro volta contengono una nota od un'annotazione. Ad alcune delle informazioni che si trovano nell'area delle note nella bibliografia convenzionale, sono stati assegnati degli elementi specifici nello schema TEI.

1f. Descrizione della fonte



L'elemento `<sourceDesc>` è un elemento obbligatorio che registra i dettagli relativi alla o alle fonti da cui è derivato il documento elettronico. Può contenere del testo libero o dei riferimenti bibliografici più o meno strutturati, mediante uno o più dei seguenti elementi:

- `<bibl>` contiene un riferimento bibliografico non rigorosamente strutturato, le cui componenti possono, o meno, essere esplicitamente marcate.
- `<biblFull>` contiene un riferimento bibliografico rigorosamente strutturato, che può contenere tutti gli elementi definiti nella TEI per la descrizione di un file.
- `<listBib>` contiene una lista di citazioni bibliografiche di qualunque tipo.

Esempi:

```
<sourceDesc>
  <bibl> Il primo folio di Shakespeare, preparato da Charlton Hinman
    (Norton fac-simile, 1968)
  </bibl>
</sourceDesc>
```

```
<sourceDesc>
  <scriptStmt id='RAI2'>
    <bibl>
      <author>RAI TG2 notizie</author>
      <title>Ultime notizie</title>
      <date>14 maggio1995</date>
    </bibl>
  </scriptStmt>
</sourceDesc>
```

2. Descrizione della codifica



L'elemento `<encodingDesc>` specifica i metodi ed i principi editoriali che hanno governato la trascrizione e la codifica di un testo. Il suo uso è fortemente consigliato. Può essere costituito da una descrizione in prosa, oppure da uno o più elementi della seguente lista:

- a. `<projectDesc>` descrive dettagliatamente il fine o lo scopo per cui un testo elettronico è stato codificato, insieme a qualunque altra informazione rilevante che riguarda il procedimento con cui il file è stato prodotto.
- b. `<samplingDecl>` contiene una descrizione dei principi e dei metodi usati nella selezione dei testi adottata per la creazione di un corpus o collezione.
- c. `<editorialDecl>` fornisce dettagli su principi e pratiche editoriali seguite nella codifica di un testo.
- d. `<tagsDecl>` fornisce informazioni dettagliate relative ai marcatori applicati a un documento XML.
- e. `<refsDecl>` specifica il modo in cui sono stati costruiti i sistemi di riferimento canonici del testo.
- f. `<classDecl>` contiene una o più tassonomie che definiscono l'insieme dei codici di classificazione impiegati nel testo.

2a/b. Descrizione del progetto e del campionamento



Esempi di `<projectDesc>` e `<samplingDecl>`:

```
<encodingDesc>
  <projectDesc>
    Testi raccolti per l'uso nella biblioteca
    dell'Università di Tor Vergata, Roma, aprile1996
  </projectDesc>
</encodingDesc>
<encodingDesc>
  <samplingDecl>
    Campioni di 2000 parole presi dall'inizio del
    testo
  </samplingDecl>
</encodingDesc>
```

2c. Dichiarazioni editoriali



L'elemento `<editorialDecl>` contiene una descrizione in prosa delle pratiche adottate nella codifica del testo. Di solito questa descrizione dovrebbe coprire i seguenti argomenti, ognuno dei quali può essere inserito in un paragrafo separato.

- **CORREZIONE** come e in quali circostanze sono state apportate correzioni nel testo.
- **NORMALIZZAZIONE** il livello di regolarizzazione o normalizzazione apportata rispetto alla fonte.
- **CITAZIONE** il trattamento riservato alle virgolette dell'originale -- se siano state conservate o sostituite da riferimenti ad entità, se siano state distinte le virgolette aperte e chiuse, etc.
- **SILLABAZIONE** il trattamento riservato ai trattini (specialmente quelli a fine riga) nell'originale -- se siano stati conservati o sostituiti da riferimenti ad entità, etc.
- **SEGMENTAZIONE** il modo in cui è stato segmentato il testo, per esempio in frasi, unità toniche, strati grafemici, etc.
- **INTERPRETAZIONE** quali informazioni analitiche o interpretative sono state aggiunte al testo.

2c. Dichiarazioni editoriali - esempio



```
<editorialDecl>
```

```
<p>La parte dell'analisi del discorso applicata  
nella sezione 4, è stata aggiunta a mano e non è  
stata controllata.
```

```
<p>Gli errori di trascrizione sono stati controllati  
utilizzando il correttore ortografico di  
WordPerfect.
```

```
<p>Tutte le parole sono state convertite in italiano  
moderno utilizzando il dizionario Garzanti.
```

```
<p>Tutte le virgolette sono state convertite in  
referenza di entità &odq; e &cdq;.
```

```
</editorialDecl>
```

2d/e/f. Dichiarazione di codifica, riferimenti e classificazioni



L'elemento `<tagsDecl>` va utilizzato per fornire informazioni dettagliate sui marcatori XML che effettivamente compaiono all'interno di un documento. Esso può contenere una semplice lista degli elementi usati, con un contatore per ognuno, composta usando i seguenti elementi speciali:

- `<tagUsage>` fornisce informazioni sull'uso di uno specifico elemento all'interno del più esterno elemento `<text>` di un documento conforme TEI. Tra gli attributi:
 - *gi* il nome (identificatore generico) dell'elemento indicato dal marcatore.
 - *occurs* specifica il numero di occorrenze di questo elemento all'interno del testo.

L'elemento `<rendition>` è usato per documentare le differenti maniere in cui gli elementi sono resi nel testo di origine.

- `<rendition>` fornisce informazioni sulla interpretazione scelta per uno o più elementi.
- `<tagUsage>` fornisce informazioni sull'uso di uno specifico elemento all'interno dell'elemento `<text>`. Tra gli attributi:
 - *occurs* specifica il numero di occorrenze di questo elemento all'interno del testo.
 - *ident* specifica il numero di occorrenze di questo elemento all'interno del testo che hanno un valore distinto dell'attributo globale *id*.
 - *render* riporta l'identificatore unico di un elemento `<rendition>` che specifica come va reso questo elemento.

Esempio di dichiarazione d'uso dei marcatori



```
<tagsDecl>
  <tagUsage gi='text' occurs='1'>
  <tagUsage gi='body' occurs='1'>
  <tagUsage gi='p' occurs='12'>
  <tagUsage gi='hi' occurs='6'>
</tagsDecl>
```

Questa dichiarazione di codifica (fittizia) sarebbe appropriata per un testo che contenga dodici paragrafi nel suo corpo, all'interno del quale sono stati marcati sei elementi `<hi>`. Da notare che se viene usato l'elemento `<tagsDecl>`, allora esso deve contenere un elemento `<tagUsage>` per ogni elemento marcato nel testo associato.

Elemento `<refsDecl>`



L'elemento `<refsDecl>` va usato per documentare il funzionamento di un eventuale schema di riferimento standard introdotto nella codifica. Nella sua forma più semplice consiste di una descrizione in prosa.

Esempio:

```
<refsDecl>
<p>L'attributo N di ogni DIV1 e DIV2 contiene il
riferimento canonica per ogni divisione nella
forma XX.YYY dove XX è il numero del libro in
numeri romani e YYY è la sezione in numeri
arabi. </p>
</refsDecl>
```

Elemento `<classDecl>`



L'elemento `<classDecl>` raggruppa le definizioni, o le fonti, di ogni schema di classificazione descrittiva usata in altre parti della testata. Almeno uno di questi schemi deve essere fornito, mediante i seguenti elementi:

`<taxonomy>`

definisce la tipologia usata per classificare testi, sia implicitamente, attraverso una citazione bibliografica, sia esplicitamente, con una tassonomia strutturata.

`<bibl>`

contiene una citazione bibliografica non strutturata le cui componenti possono o meno essere etichettate esplicitamente.

`<category>`

contiene una singola categoria descrittiva, eventualmente inserita all'interno di una categoria di livello superiore, in una tassonomia definita dall'utente.

`<catDesc>`

descrive una data categoria nell'ambito di una tassonomia o di una tipologia testuale, sotto forma di una breve descrizione in prosa.

Esempi



Nel più semplice dei casi, la tassonomia può essere definita da un riferimento bibliografico:

```
<classDecl>
  <taxonomy id='LAI'>
    <bibl>Libreria dell'area di italianistica</bibl>
  </taxonomy>
</classDecl>
```

Altrimenti, o in aggiunta, il codificatore può definire uno speciale schema di classificazione mirata:

```
<taxonomy id=B>
<bibl>Brown Corpus</bibl>
  <category id='B.A'><catDesc>Servizio di stampa
  <category id='B.A1'><catDesc>Quotidiano</category>
  <category id='B.A2'><catDesc>Settimanale</category>
  <category id='B.A3'><catDesc>Nazionale</category>
  <category id='B.A4'><catDesc>Provinciale</category>
  <category id='B.A5'><catDesc>Politico</category>
  <category id='B.A6'><catDesc>Sportivo</category>
  ...
  <category id='B.D'><catDesc>Religione
  <category id='B.D1'><catDesc>Libri</category>
  <category id='B.D2'><catDesc>Periodici e opuscoli</category>
</category>
...
</taxonomy>
```

Il collegamento tra un testo particolare ed una categoria all'interno di tale tassonomia viene realizzato mediante l'elemento `<catRef>`, inserito nell'elemento `<textClass>`, come ulteriormente descritto più avanti.

3. Descrizione del profilo



L'elemento `<profileDesc>` permette di registrare, in un quadro unico, una serie di informazioni che caratterizzano, sotto vari aspetti descrittivi, un testo. Esso ha tre componenti opzionali:

`<creation>`

contiene informazioni relative alla creazione di un testo.

`<langUsage>`

descrive le lingue, sottolingue, registri, dialetti, etc., rappresentati all'interno di un testo.

`<textClass>`

contiene informazioni che descrivono la natura o i temi di un testo, in termini di uno schema di classificazione standard, di un thesaurus, etc.

Esempio di utilizzo dell'elemento `<creation>`:

```
<creation>
  <date value='8-1992'>Agosto 1992</date>
  <name type='luogo'>Mantova, Italia</name>
</creation>
```

Elemento <textClass>



L'elemento *<textClass>* classifica un testo in base al sistema o ai sistemi definiti nell'elemento *<classDecl>*, e contiene uno, o più, dei seguenti elementi:

- *<keywords>* contiene una lista di parole chiave o di espressioni che identificano il tema o la natura di un testo. Tra gli attributi:
 - *scheme* identifica il vocabolario controllato all'interno del quale è definito l'insieme di parole chiave.
- *<classCode>* contiene il codice di classificazione usato per il testo in un sistema di classificazione standard. Tra gli attributi:
 - *scheme* identifica il sistema di classificazione o la tassonomia in uso.
- *<catRef>* specifica una, o più categorie definite nell'ambito di una data tassonomia o tipologia di testi. Tra gli attributi:
 - *target* identifica le categorie in questione.

Elemento <keywords>



L'elemento *<keywords>* contiene una lista di parole chiave o espressioni che identificano il tema o la natura del testo. L'attributo *scheme* le collega al sistema di classificazione definito in *<taxonomy>*.

```
<textClass>
  <keywords scheme='LAI'>
    <list>
      <item>Letteratura italiana--Storia e critica--Data
dell'elaborazione</item>
      <item>Letteratura italiana--Storia e critica--
Teoria, etc.</item>
      <item>Lingua italiana--Stile--Data
dell'elaborazione.</item>
    </list>
  </keywords>
</textClass>
```

4. Descrizione delle revisioni



L'elemento `<revisionDesc>` fornisce un "diario" dei cambiamenti in cui può essere registrato ogni cambiamento apportato ad un testo.

Il "diario" può essere registrato come una sequenza di elementi `<change>` ognuno dei quali contiene:

`<date>` contiene una data in qualunque formato.

`<respStmt>` fornisce una dichiarazione di responsabilità relativa al responsabile del contenuto intellettuale di un testo, un'edizione, una registrazione, una collana, se gli elementi specifici per autori, curatori, etc., non sono sufficienti o non sono adatti.

`<item>` contiene un componente di una lista.

Esempio:

```
<revisionDesc>
  <change><date 3/6/1991:</date>
  <respStmt><name>EMB</name><resp>ed.</resp></respStmt>
  <item>Formato del file sopradatato</item>
  <change><date>25/5/1990:</date>
  <respStmt><name>EMB</name><resp>ed.</resp>
  <item>correzioni registrate da Stuart</item>
</revisionDesc>
```