

I linguaggi di markup e la rappresentazione (complessa) di testi complessi

Fabio Ciotti

La complessità del testo

- Il testo è un oggetto complesso in quanto è in grado di veicolare significato (e di rivestire dunque interesse scientifico) su più livelli strutturali, e cosa ancora più rilevante attraverso l'instaurazione di molteplici relazioni tra più livelli

La complessità del testo

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

La complessità del testo

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

La complessità del testo

- XML (e SGML) può essere considerato sia un formalismo sia un modello di dati espresso da quel formalismo, e tale (meta)modello è appunto un albero ordinato etichettato.
- XML considerato come linguaggio di modellizzazione può esprimere solo modelli la cui struttura è un albero
- Tutte le applicazioni XML standard sono in grado di elaborare (validare, trasformare, formattare) strutture dati ad albero

Il problema delle gerarchie sovrapposte

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

Il problema delle gerarchie sovrapposte

- compresenza di due o più strutture (livelli) gerarchiche i cui elementi si sovrappongono

SCENA III

ARGIA, ANTIGONE

Argia

Una infelice io sono.

Antigone

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

Argia

Io... cerco...

... d'Antigone...

Antigone

Perché? — Ma tu, chi sei?

Antigone conosci? a lei se' nota?

che hai seco a far? che hai tu comun con essa?

Il problema delle gerarchie sovrapposte

- un elemento appartenente a una gerarchia che si estende oltre i confini dell'elemento in cui inizia

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

Il problema delle gerarchie sovrapposte

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

Il problema delle gerarchie sovrapposte

- Elementi che possono auto-sovrapporsi illimitatamente.
 - lasciato parte della sua gente {(<alla [defesa) della> obsidione]}, con le
 - lasciato parte della sua gente <alla offesa delle> obsidione, con le
 - lasciato parte della sua gente alla [difesa di quella], con le
 - lasciato parte della sua gente (difesa) della obsidione, con le
 - gente {allobsidione}, con le

Soluzioni per il problema OH

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

Soluzioni per il problema OH

- Human readability
- Maintainability
- Available implementations
- XML compatibility
- Ease of validation
- Validation across hierarchies
- Ease of formatting
- Ease of extracting multiple views
- Ease of extracting hierarchical subsets
- Continuity of text content [De Rose 2004]

Soluzioni per il problema OH

- Segmentazione

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

Soluzioni per il problema OH

- Elementi join

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

...

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

Soluzioni per il problema OH

- Stand-off markup

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

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

Soluzioni per il problema OH

- Elementi milestone

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

Soluzioni per il problema OH

- Elementi milestone

```
lasciato parte della sua gente  
<anchor id=p1/>alla  
<anchor id=p2/>defesa  
<anchor id=p3/>della  
<anchor id=p4/>obsidione  
<anchor id=p5/>, con le  
<-- ... -->
```

```
<app from=p1 to=p4>  
<lemma wit="G L P W M U">alla difesa della</>  
<rdg wit="D">alla offesa delle</>  
</app>
```

```
<app from=p2 to=p5>  
<lemma wit="G L P W M U">defesa della obsidione</>  
<rdg wit="R.Bla">difesa di quella</>  
</app>
```

```
<app from=p1 to=p3>  
<lemma wit="G L P W M U">alla difesa</>  
<rdg wit="B">difesa</>  
</app>
```

```
<app from=p1 to=p5>  
<lemma wit="G L P W M U">alla difesa della obsidione</>  
<rdg wit="A E">allobsidione</>  
</app>
```

Soluzioni per il problema OH

- XConcur

```
<?xconcur version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE (I1)div SYSTEM "tei/dtd/teispok2.dtd">
<!DOCTYPE (I2)text SYSTEM "tei/dtd/teiana2.dtd">
<(I1)div type="dialog" org="uniform">
  <(I2)text>
    <(I1)u who="Peter">
      <(I2)s>Hey Paul!</(I2)s>
      <(I2)s>Would you give me
    </(I1)u>
    <(I1)u who="Paul">the hammer?
      </(I2)s>
    </(I1)u>
  </(I2)text>
</(I1)div>
```

- JITTs

Soluzioni per il problema OH

- Layered Markup and Annotation Language (LMNL)
 - range
 - Layer

[l]{w}One{w} [w]is{w} [w]be-{l}

[l]fore{w} [w]two{w}.{l}

Soluzioni per il problema OH

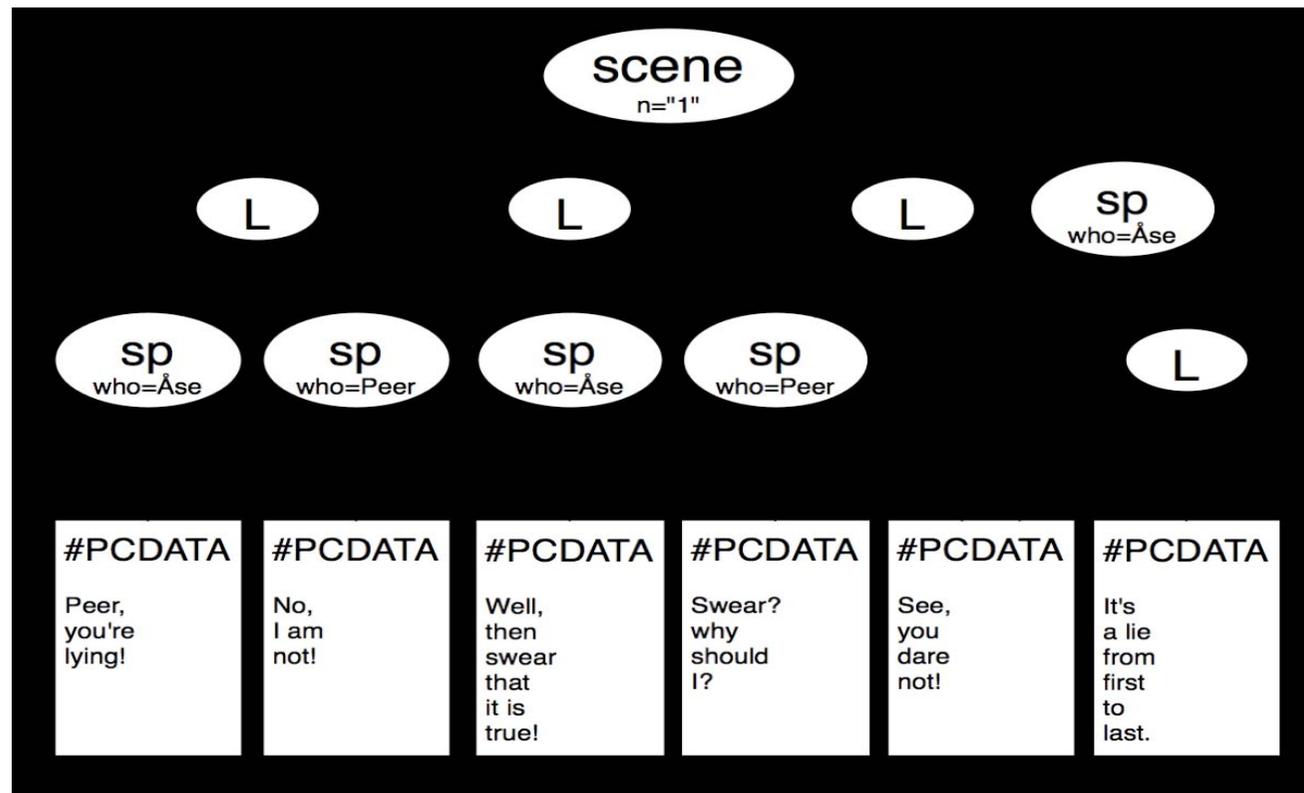
- TexMECS e GODDAD

- Notazione: TexMECS

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

Soluzioni per il problema OH

- Data model: GODDAD (*Generalized Ordered-Descendant Directed Acyclic Graph*)



Conclusioni

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