



HAL
open science

Modelling with the TEI

Marta Materni

► **To cite this version:**

Marta Materni. Modelling with the TEI. École thématique. EDEEN École d'Été en Édition Numérique, France. 2018. cel-02073430

HAL Id: cel-02073430

<https://hal.science/cel-02073430v1>

Submitted on 19 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The logo for 'Eddeen' is rendered in a large, white, stylized font with a 3D effect. It is set against a background collage that includes historical documents with circular stamps, a pink flower, and a green-tinted area with faint HTML code snippets like '</tr>', '</table>', and '</div>'.

Modelling with the TEI

Marta Materni
Université Grenoble Alpes
Marie Curie Fellow

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 745821 (DigiFlor)

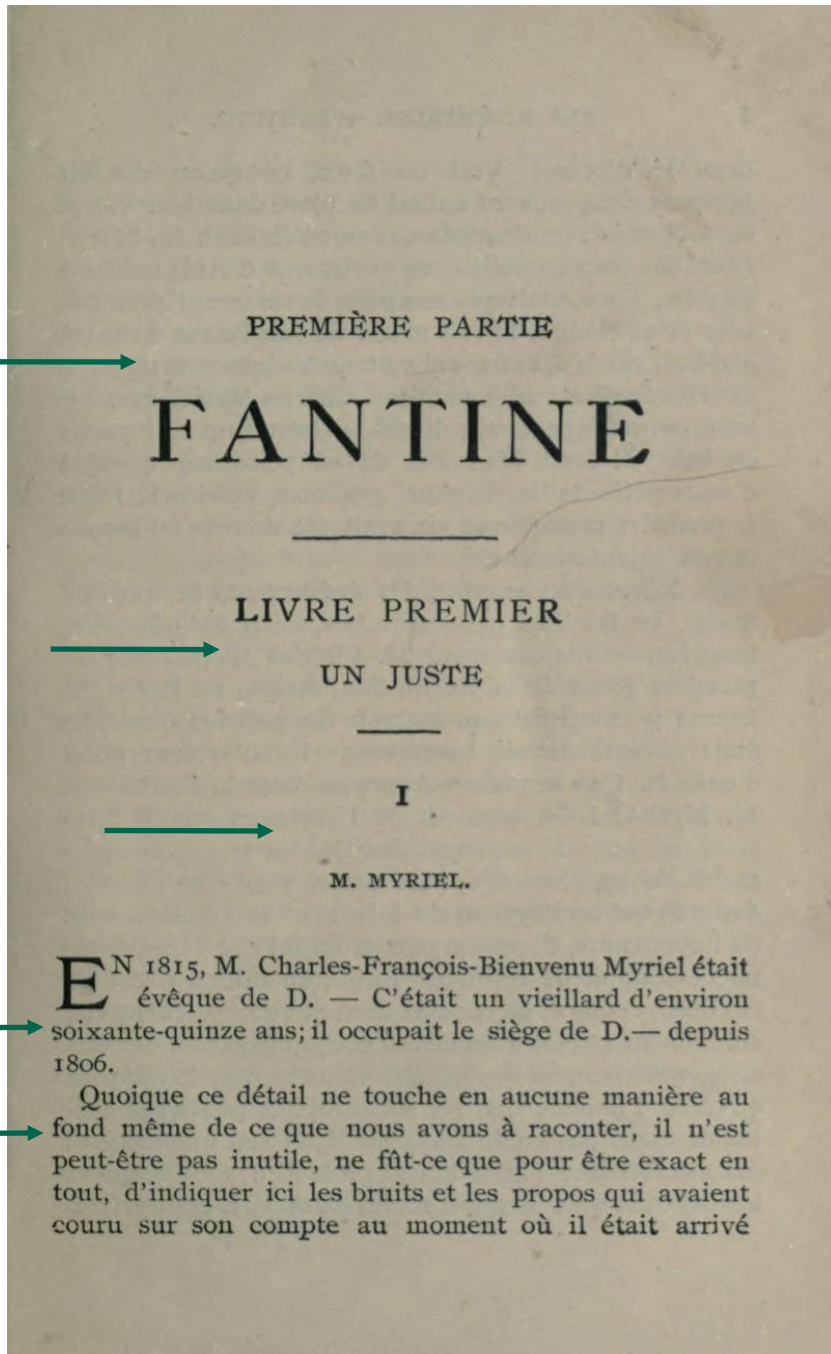
1 The Structure of the Text

First (and fundamental) question :

Is my text *prose* or *verse* ?

Second (no less fundamental) question :

Is my text a *Novel*,
an *Essay*,
a *Poem*,
a *Drama*,
a *Letter*,
an *Archival Document* ?



The Prose (Novel, Essay, etc.)

Part

Book

Chapter

Subchapter

Paragraph

The latter is the core of a prose text.

It contains a semantic unit.

The layout and the typographic code indicates it via a new line and by justifying the right margin.

Prose Hierarchy & TEI

Part / Book / Chapter / Subchapter :

`<div>` “text division” + `@type` (eventually `@n`)

BUT the characterizing and distinctive element is :

`<p>` (paragraph)

It's contained by: the `<body>` or a `<div>`

It contains: everything up to the sentence-level
(`<s>`: sentence, `<w>`: word, `<c>`: character)

Les Misérables

Première Partie – Fantine

Livre Premier – Un juste

I – M. Myriel

En 1815, M. Charles-François-Bienvenue Myriel était évêque de D. - C'était un vieillard d'environ soixante-quinze ans ; il occupait le siège de D. - depuis 1806...

```

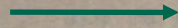
<text>
  < body>
    <div type="partie" n="1">
      <head>Première Partie – Fantine</head>

      <div type="livre" n="1">
        <head>Livre Premier – Un juste</head>

        <div type="chapitre" n="1">
          <head>I. M. Myriel</head>

          <p>En 1815, M. Charles-François-Bienvenue
            Myriel était évêque de D. - C'était un vieillard
            d'environ soixante-quinze ans ; il occupait le siège de D. - depuis 1806...
          </p>
        </div>
      </div>
    </div>
  </body>
</text>


```





III



ÉLÉVATION



Au-dessus des étangs, au-dessus des vallées,
Des montagnes, des bois, des nuages, des mers,
Par-delà le soleil, par-delà les éthers,
Par-delà les confins des sphères étoilées,



Mon esprit, tu te meus avec agilité,
Et, comme un bon nageur qui se pâme dans l'onde,
Tu sillones galment l'immensité profonde
Avec une indicible et mâle volupté.

The Poem

Section within a collection

Poem

Structure of the poem

Verse

The latter is the core of a poem:
a poem is a Text in Verse.

It has usually (but not always)
a clearly formalized layout.

Poem Hierarchy & TEI

Every poem is an independant text within macrotext/volume

Volume / Section :

<div> + **@type** (eventually **@n**)

BUT the characterizing and distinctive elements are :

<l> (line)
<lg> (line group)
+ **@type**, **@n**

Les fleurs du mal

Spleen et idéal

III. Élévation

Au-dessus des étangs, au-dessus des vallées,
Des montagnes, des bois, des nuages, des mers,
Par-delà le soleil, par-delà les éthers
Par-delà les confins des sphères étoilées

Mon esprit, tu te meurs avec agilité,
etc.

```
<text>
<body>
  <head>Les fleurs du mal</head>
```

```
<div type="sélection" n="1">
  <head>Spleen et idéal</head>
```

```
<div type="sonnet » n="3">
  <head>Élévation</head>
```

```
<lg n="1" type="quatrain">
```

```
<l>Au-dessus des étangs, au-dessus des vallées,</l>
```

```
<l>Des montagnes, des bois, des nuages, des mers,</l>
```

```
<l>Par-delà le soleil, par-delà les éthers</l>
```

```
<l>Par-delà les confins des sphères étoilées</l>
```

```
</lg>
```

```
<lg n="2" type="quatrain">
```

```
<l>Mon esprit, tu te meurs avec agilité,</l>
```

```
etc.
```

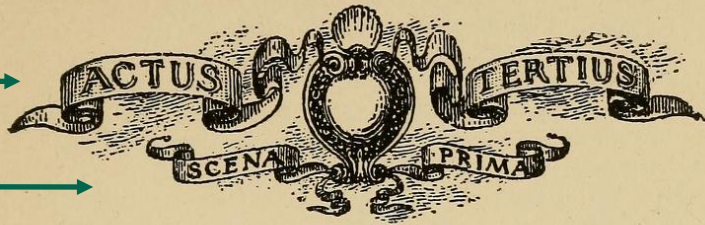
```
</lg>
```

```
</div>
```

```
</div>
```

```
</body>
```

```
</text>
```



The Capitol; the Senate sitting.

A crowd of People in the Street leading to the Capitol; among them ARTEMIDORUS and the Soothsayer. Flourish. Enter CÆSAR, BRUTUS, CASSIUS, CASCA, DECIUS, METELLUS, TREBONIUS, CINNA, ANTONY, LEPIDUS, POPILIUS, PUBLIUS, and others.

Cæsar. The ides of March are come.

Soothsayer. Ay, Cæsar; but not gone.

Artemidorus. Hail, Cæsar! Read this schedule.

Decius. Trebonius doth desire you to o'er-read,
At your best leisure,* this his humble suit.

*Cf. Bacon: "Yf yow he at leasure."—*Prom. No.* 1375 (1594).

"If your lordship were at leisure."—*Ham. V.*, 2.

"If your leisure served."—*M. Ado. III.*, 2.

"Had you such leisure."—*R. III.*, 1., 2.

Artemidorus. O, Cæsar, read mine first; for mine's a suit

That touches Cæsar nearer. Read it, great Cæsar.

Cæsar. What touches us ourself shall be last serv'd.

Artemidorus. Delay not, Cæsar; read it instantly.

Cæsar. What! is the fellow mad?

Publius. Sirrah, give place.

Cassius. What! urge you your petitions in the street?

Come to the Capitol.

CÆSAR enters the Capitol, the rest following. All the Senators rise.

Drama

Act

Scene

Setting

Speakers

Text

A very stable and clearly formalized structure.

A peculiar division of the text : Act, Scene

The text could be both in prose and in verse : the structure is therefore its really distinctive feature.

Drama Hierarchy & TEI

Act / Scene : `<div>` + `@type` (eventually `@n`)

BUT the characterizing and distinctive elements are :

`<stage>`

(stage direction) « any kind of stage direction within a dramatic text » + `@type` (*setting, entrance, exit, novelistic, etc.*)

`<sp>`

(speech) « an individual speech in a performance text »

`<speaker>`

« a specialized form of heading or label, giving the name of one or more speakers in a dramatic text »

And then: the text as `<l>` or `<p>` ... and all their encoding rules

Act III
Scene II.

The same. The Forum.
Enter BRUTUS and CASSIUS and a trough of
CITIZENS.

MARCUS ANTONIUS.
You gentle Romans, ...

CITIZENS.
Peace, ho ! Let us hear him.

MARCUS ANTONIUS.
Friends, Romans, countrymen, lend me your ears ;
I Come to bury Caesar, not to praise him.
Here, under leave of Brutus and the rest, -
For Brutus is an honourable man ;
So are they all, all honourable men, -
Come I to speal in Caesar's funeral.
He was my friend, faithful and just to me :
But Brutur says he was ambitious ;
And Brutus is an honourable man,

```
<text>
<body>
  <div type="act" n="3">
    <head>Act III</head>
    <div type="scene" n="2">
      <head>Scene II</head>
      <stage type="location">The same. The Forum.</stage>
      <stage type="entrance">Enter BRUTUS and CASSIUS and a
trough of CITIZENS.</stage>
      <sp>
        <speaker>MARCUS ANTONIUS.</speaker>
        <l>You gentle Romans, ...</l>
      </sp>
      <sp>
        <speaker>CITIZENS.</speaker>
        <l>Peace, ho ! Let us hear him.</l>
      </sp>
      <sp>
        <speaker>MARCUS ANTONIUS.</speaker>
        <l>Friends, Romans, countrymen, lend me your ears ;</l>
        <l>I Come to bury Caesar, not to praise him.</l>
        <l>Here, under leave of Brutus and the rest, -</l>
        <l>For Brutus is an honourable man ;</l>
        <l>So are they all, all honourable men.</l>
      </sp>
    </div>
  </div>
</body>
</text>
```

LETTRES

CHOISIES

DE MADAME DE SEVIGNE.

(39.) *A madame* DE GRIGNAN.

Aux Rochers, mercredi 15 juillet 1671.

Si je vous écrivais toutes mes rêveries sur votre sujet, je vous écrirais toujours les plus grandes lettres du monde; mais cela n'est pas bien aisé: ainsi je me contente de ce qui peut s'écrire, et je rêve tout ce qui peut se rêver: j'en ai le temps et le lieu. La Mousse a une petite fluxion sur les dents, et l'abbé a une petite fluxion sur le genou, qui me laissent le champ libre dans mon mail, pour y faire tout ce qui me plaît. Il me plaît de m'y promener le soir jusqu'à huit heures; mon fils n'y est plus; cela fait un silence, une tranquillité et une solitude que je ne crois pas qu'il soit aisé de rencontrer ailleurs. Je ne vous dis point à qui je

I.

The Letter

Section of the collection

Opener

Body of the letter

Closer

Real or fictive, prose or verse: the key is in its *structure* and *function*, that means the communication between two people.

It has a clearly formalized layout.

Dates and places, opening and closing formulas: the core of a letter (after the text, obviously!)

Letter Hierarchy & TEI

Section / Letter : `<div>` + `@type` (eventually `@n`)

The text : `<p>` (or `<l>`)... and all their encoding rules.

BUT MANY characterizing and distinctive elements :

`<opener>`

« groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter »

`<closer>`

« groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter »

`<postscript>`

« contains a postscript, e.g. to a letter »

And MORE in detail :

<byline>

« contains the primary statement of responsibility given for a work on its title page or at the head or end of the work »

<dateline>

« contains a brief description of the place, date, time, etc. of production of a letter » (+ <date>, <place> etc.)

<salute>

« contains a salutation or greeting prefixed to a foreword, dedicatory epistle, or the salutation in the closing of a letter »

<address>

« contains a postal address »

<signed>

« contains the closing salutation, etc., appended to a foreword, dedicatory epistle ».

```
<div type="letter-poem">
```

```
<opener>
```

```
<dateline><place>Assemblée nationale,</place></dateline>
```

```
<address>
```

```
<addrLine>à Mademoiselle Anne Pingeot,</addrLine>
```

```
<addrLine>39 rue de Cherche-Midi,</addrLine>
```

```
<addrLine>Paris VIe.</addrLine>
```

```
</address>
```

```
<dateline><date>Mardi 1er décembre 1964</date></dateline>
```

```
</opener>
```

```
<p> <lb/>Que ces roses, mon amour, soient pour toi le reflet des beaux jours que nous venons  
de vivre; <lb/>Qu'elles soient le prélude d'un mois de joie et de ferveur, ces roses de décembre;  
<lb/>Qu'elles soient enfin le signe d'un coeur en paix - <lb/>Et le mien le sera <lb/>Si tu vieux  
bien, d'un sourire, <lb/>Me dire que je suis pardonné.</p>
```

```
<closer>
```

```
<salute>Je t'embrasse et je t'aime</salute>
```

```
<signed>François</signed>
```

```
</closer>
```

```
<postscript><p>ps. Je serai rue du Regard à 13 h 15 et nous irons déjeuner ensemble, ma  
chérie.</p></postscript>
```

```
</div>
```

2 Inside the Text

Highlighting : `<hi>` & the others...

Encoding —————→ WYSIWYM (What You See Is What You Mean)
and not *Copying* —————→ WYSIWYG (What You See is What You Get)

In the world of Typography: *italic* or “...”

But in the world of Encoding?

From Implicit to Explicit.

- 1 Le monde est le théâtre de la comédie humaine.
- 2 La « comédie humaine » de la ville bourgeoise.
- 3 La *Comédie Humaine* est composée par 137 volumes.

- 1 `<p>`Le monde est le théâtre de la comédie humaine`</p>`
- 2 `<p>`La `<soCalled>`comédie humaine`</soCalled>` de la ville bourgeoise`</p>`
- 3 `<p>`La `<title>`Comédie Humaine`</title>` est composée par 137 volumes`</p>`

A generic encoding :

<hi> (highlighted)

« marks a word or phrase as *graphically distinct* from the surrounding text »

A more detailed encoding :

<distinct>

« identifies any word or phrase which is regarded as *linguistically distinct*, for example as archaic, technical, dialectal, etc. »

<foreign>

« identifies a word or phrase as belonging to some language other than of the surrounding text »

<emph>

« marks words or phrases which are stressed or emphasized for linguistic and rhetorical effect »

<mentioned>

« marks words or phrases mentioned, not used »

<term>

« contains a word which is regarded as a technical term » etc. etc. etc.

It's not a problem of "physical appearance" :

Great Expectations is Dickens' masterpiece
« Great Expectations » is Dickens' masterpiece
Great Expectations is Dickens' masterpiece

What does it *means* Great Expectations inside my text?

It's a TITLE

typographically the TITLES are marked by *italic*, « » , etc.

but their essence is to be a TITLE

variously HIGHLIGHTED according to a style

Its copy WYSIWYG is:

Great Expectations, « Great Expectations », Great Expectations

Its encoding WYSIWYM is:

<title>Great Expectations</title>

Quotations : <said> or <quote> ?

1 - Mais tu va pleurer !, dit le petit prince.

- Bien sur, dit le renard.

- Alors tu n'y gagnes rien !

- J'y gagne, dit le renard, à cause de la couleur du blé.

2 Souvent j'ai dit à mes élèves qu'il en vaut la peine de gagner
« à cause de la couleur du blé ».

1 <said>Mais tu va pleurer !</said>, dit le petit prince.

<said>Bien sur</said>, dit le renard.

<said>Alors tu n'y gagnes rien !</said>

<said>J'y gagne</said>, dit le renard, <said>à cause de la couleur du blé</said>.

2 Souvent j'ai dit à mes élèves qu'il en vaut la peine de gagner
<quote>à cause de la couleur du blé</quote>.

<said>

(speech or thought)

« indicates passages thought or spoken aloud, whether explicitly indicated in the source or not, whether directly or indirectly reported, whether by real people or fictional characters »

<quote>

(quotation)

« contains a phrase or passage attributed by the narrator or author to some agency *external* to the text »

Named Entities : Places & People

The most generic encoding :

`<rs>`

(referencing string)

« contains a general purpose name or referring string »

+ `@type` (person, place, etc.)

A little less generic :

`<name>`

(name, proper noun)

« contains a proper noun or noun phrase »

+ `@type` (person, place, etc.)

1 The `<rs>`Queen of the United Kingdom`</rs>`.

2 `<name>`Elizabeth II`</name>` is the Queen of the United Kingdom.

A sentence from *Pride and Prejudice* in the Guidelines :

My dear **<rs type="person" ref="#BE">**Mr. Bennet**</rs>**,
said **<rs type="person" ref="#MI">**his lady**</rs>** to him one day,
have you heard that
<rs type="place" ref="#NP">Netherfield Park**</rs>**
is let at least ?

Or (it's equivalent) :

My dear **<name type="person" ref="#Ben">**Mr. Bennet**</rs>**,
said **<rs type="person" ref="#MI">**his lady**</rs>** to him one day,
have you heard that
<name type="place" ref="#NP">Netherfield Park**</name>**
is let at least ?

A very detailed encoding :

```
<persName> : Charles de Gaulle  
  <placeName> : Paris  
<geogName> : Mediterranean Sea
```

A mixed encoding, for even more detail :

```
the <placeName ref="#PE" >Pilars of  
  <persName ref="#HE">Hercules</persName></placeName>
```

The importance of *a descriptive key* :

@ref « reference » but... to what ?

To a detailed description of people, places, etc.

WHAT : a `<listPerson>` or a `<listPlace>`

WHERE : inside the `<teiHeader>` or inside a `<div>/<p>`
(inside the `<text>`, maybe in the `<back>`).

```
<person xml:id="SE">
  <persName xml:lang="fr">
    <forename>Antoine</forename>
    <forename>Jean Baptiste</forename>
    <forename>Marie</forename>
    <forename>Roger</forename>
    <roleName type="nobility">conte</roleName>
    <nameLink>de</nameLink>
    <surname>Saint-Exupéry</surname>
    <addName>Saint-Ex</addName>
    <roleName type="honorific">Officier de la Légion d'honneur</roleName>
    <roleName type="honorific">Croix de guerre</roleName>
  </persName>
  <birth>
    <date when="1900-06-29">29 Juin 1900</date>
    <placeName>Lyon</placeName>
  </birth>
  <death>
    <date when="1944-07-31">31 Juillet 1944</date>
    <geogName>Golfe de Marseille</geogName>
  </death>
</person>
```

Time

A generic date :

<date>

« contains a date in any format »

A moment in time :

<time>

« contains a phrase defining a time of day in any format »

It's important to use **@when** for a standardised (and therefore searchable) expression of the time :

yyyy-mm-dd for <date>

00:00:00 for <time>

On **<date when="2018-05-02">**the second day of May**</date>**,
at **<time when="08:00:00">**eight o'clock in the morning**</time>**,
the inscriptions were opened

The complexity of the time :

The year 2018 :

`when="2018 »`

May 2018 :

`when="2018-05 »`

28 May 2018 :

`when="2018-05-28 »`

May :

`when="—05 »`

The 28th day :

`when="---28 »`

From 28 May to 2 June 2018 :

`from="2018-05-28" to="2018-06-02 »`

The third week of May 2018 :

`notBefore="2018-05-14" notAfter="2018-05-21"`

Exercises: folder Prose/Dickens, file ex_dickens, ex. 3
folder Letter/Sevigne, file ex_sevigne, ex. 2

Conclusions

first, the big divide :

It is my text a *prose* or a *poem*?

second :

What is the *literary genre* of my text?

then, inside the text:

What do I want *to explicit*?

- Highlights
- Quotations
- People and Places
- Time
- etc. etc. etc.....