How to use the TEI Guidelines

Five steps on the road to TEI enlightenment

1. Modelling : what sort of textual objects are you trying to represent ?
2. Orientation : are these objects already known to the TEI ?
3. Declaration : how do you make a TEI-conformant schema for them ?
4. Documentation : how do you document your encoding practice ?
5. Validation : how do you check that all this works as it should ?

Modelling

Modelling your data is an essential first step. Whether you plan to use UML, a RDBMS, SKOS, RDF or whatever, if you don't start from a formal and explicit model of the things your information system is intended to handle, you will find it very difficult to build.

Traditional data analysis/modelling

First identify ..

- The "objects of interest"
- Their properties (attributes)
- Relationships between those objects
- The processes that will be performed on or with those objects

This analysis is needed whether you use the TEI or not!

How to find you way in the TEI

The TEI Guidelines, printed or online, is divided into two parts.

- 23 chapters, most of them defining a 'module', i.e. a group of elements and attributes
- alphabetical reference lists defining classes (190), elements (602), attributes (506), macros (10) and datatypes (28)

How do you decide which TEI element (etc) you will use to model which object in your conceptual model? How do you find out that you will need a ‘unicorn’ ?

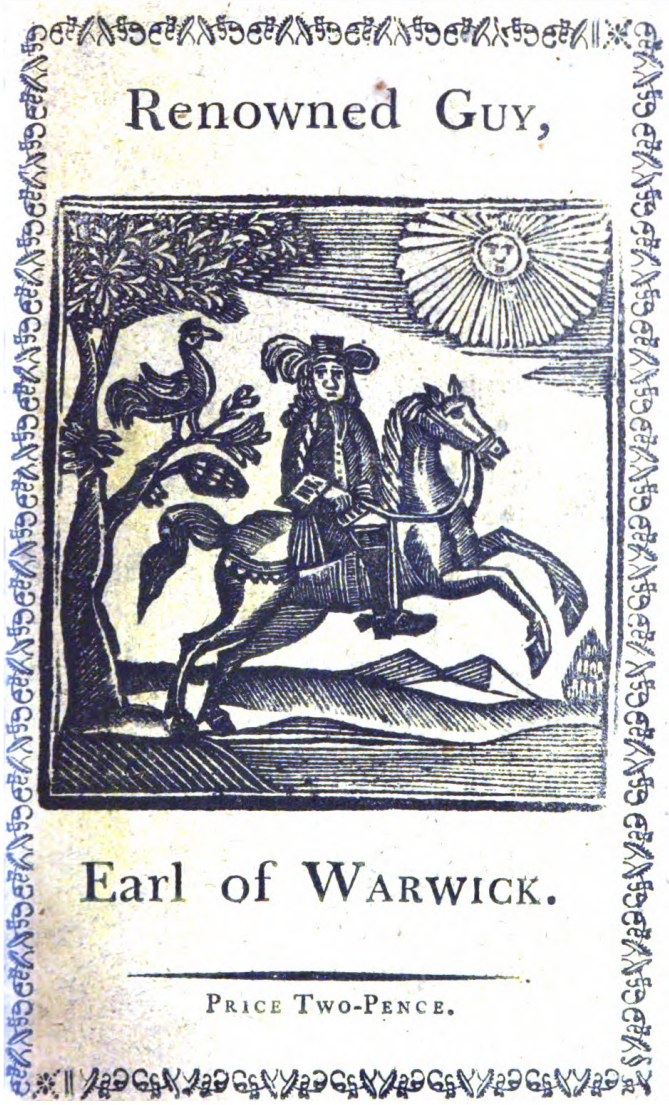
The sad truth ..

- Sadly there's no short cut answer to such questions
- You have to look carefully at the examples and the documentation to see whether the TEI's idea of what a ‘blort’ is corresponds with yours
- (though work in the Ontology SIG does demonstrate that much of the TEI conceptual model can be represented as an ontology)

For each object in your model, you must decide :

- which existing TEI object corresponds exactly with it (if any)
- which existing TEI object is close enough, and what modification would be needed to make it better?
- what objects in your analysis are entirely missinmg from the TEI ? (yes, unicorns do exist!)

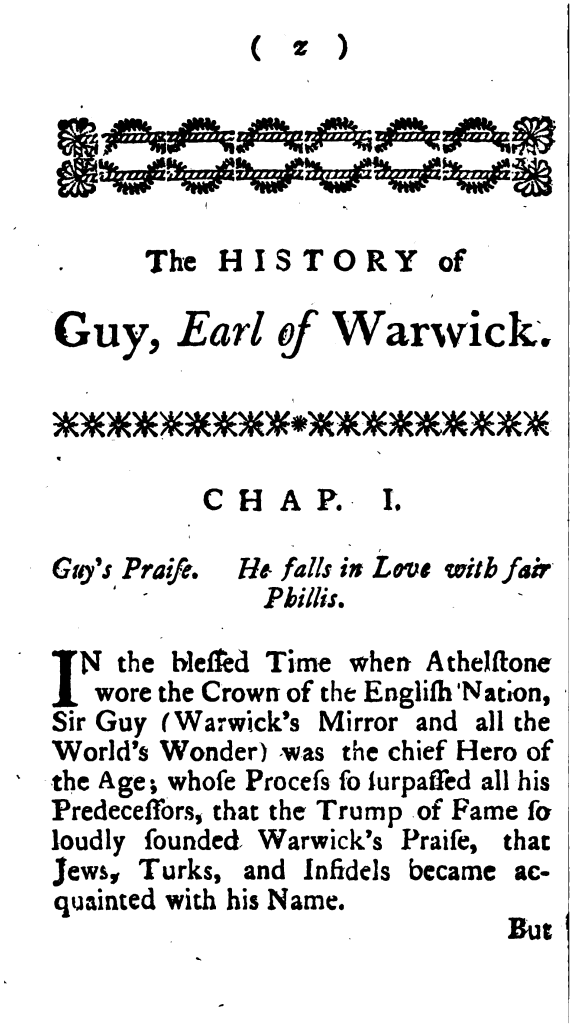
A simple example



Structure

- It's a book (a chapbook) with a traditional structure (titlepage, chapters, etc.)
- Each chapter contains :
  - a heading and a brief summary
  - paragraphs of narrative
  - ostensible passages of dialogue
  - Names of people, places, events
- Many other versions of the same narrative exist in many other books
- Each page has paratextual features (running heads, signature, page numbers)
- And there is much ornamentation and illustration

For example...



(z)

The HISTORY of

Guy, *Earl of* Warwick.

CHAP. I.

Guy's Praiſe. He falls in Love with fair Phillis.

IN the bleſſed Time when Athelſtone wore the Crown of the Engliſh Nation, Sir Guy (Warwick's Mirror and all the World's Wonder) was the chief Hero of the Age; whoſe Proceſs ſo ſurpaſſed all his Predeceſſors, that the Trump of Fame ſo loudly ſounded Warwick's Praiſe, that Jews, Turks, and Infidels became acquainted with his Name.

But

TEI document structure(s)

- A TEI document is represented by means of a <TEI> element, which contains both data and metadata
- A group of <TEI> elements with shared metadata can be combined to form a <teiCorpus>

We will treat this chapbook as a single object, and therefore as a single<TEI> element. For the moment, we're concerned just to capture its text, though there are other possibilities

TEI basic structure (1)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- required : contains metadata -->
 </teiHeader>
 <facsimile>
<!-- optional : collection of images of the object-->
 </facsimile>
 <sourceDoc>
<!-- optional : documentary transcription of the object -->
 </sourceDoc>
 <text>
<!-- required if no facsimile or sourceDoc -->
 </text>
</TEI>
```

TEI basic structure (2)

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- required -->
 </teiHeader>
 <TEI>
<!-- required -->
 </TEI>
<!-- More <TEI> elements -->
</teiCorpus>
```

We might use this for a collection of chapbooks, or for a curated set of versions of the text

The <text> element

What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter

The <text> element

What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter

The <text> element

What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter

TEI text structure (1)

A simple document:

```
<text>
 <front>
<!-- optional -->
 </front>
 <body>
<!-- required -->
 </body>
 <back>
<!-- optional -->
 </back>
</text>
```

Composite texts

There are also composite texts (eg collected poems, a book of essays), which contain

- optional front matter
- a <group> with <text> inside (required)
- optional back matter

TEI text structure (2)

```
<text>
 <front>
<!-- ... -->
 </front>
 <group>
  <text n="1">
   <body>
    <p>...</p>
   </body>
  </text>
  <text n="2">
   <body>
    <p>...</p>
   </body>
  </text>
 </group>
 <back>
<!-- ... -->
 </back>
</text>
```

What's inside the <body> of a <text>...

Usually, there are divisions of various kinds

- which may nest one with another
- which have a @type for example "chapter", "subsection" etc.
- and possibly a name or number of some kind (for which we use the @n attribute)

Two other useful attributes:

- @xml:id provides a unique identifier
- @xml:lang identifies the human language of the content of the division

These are two of the socalled global attributes, which can be specified for any TEI element

Global attributes

Some features (potentially) apply to everything:

- identity
- language
- rendition

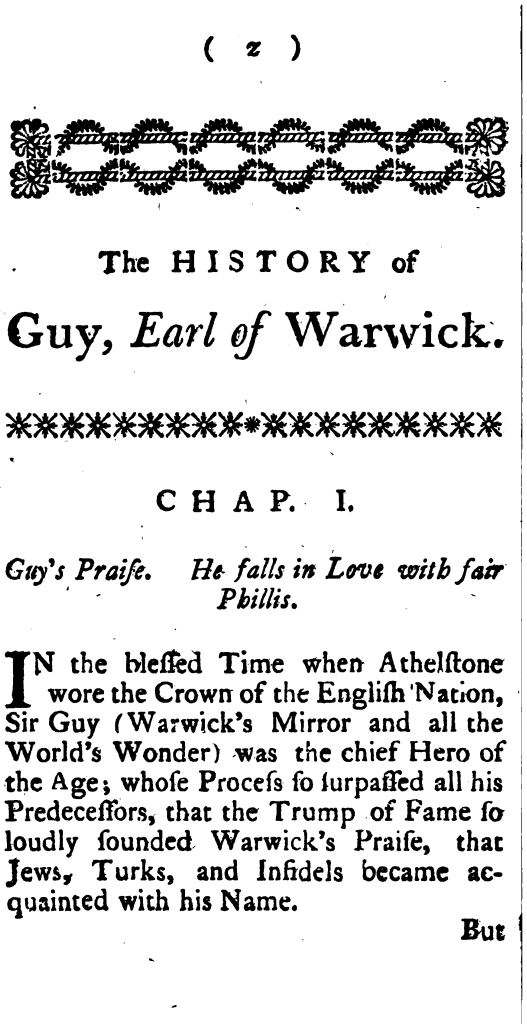
TEI provides global attributes for these:

- @xml:id provides a unique identifier for any element;
- @n provides a name or number for any element
- @xml:lang specifies the language of any element, using an ISO standard code
- @rend, @style, and @rendition provide ways of specifying the visual appearance (rendition) of any element

An example ...

```
<body xml:lang="fr">
 <head>Les Chouans</head>
 <div type="partie" n="I" xml:id="CHOUAN0100">
  <head>Première partie</head>
  <head>L'embuscade</head>
  <div n="1" type="chapitre" xml:id="CHOUAN0101">
   <head>Chapitre premier</head>
   <p>Dans les premiers jours de l'an VIII, au commencement de
vendémiaire, ou, pour se
        conformer au calendrier actuel, vers la fin du mois de
septembre 1799... </p>
<!-- suite du chapitre 1 -->
  </div>
  <div n="2" type="chapitre" xml:id="CHOUAN0102">
   <head>Chapitre deuxieme</head>
<!-- chapitre 2 -->
  </div>
<!-- suite du premier livre -->
  <trailer> Fin de la premiere partie</trailer>
 </div>
 <div type="partie" n="II" xml:id="CHOUAN0100">
  <head> Deuxieme partie</head>
<!-- deuxieme livre-->
 </div>
<!-- d'autres livres -->
</body>
```

Or, in our chapbook....



(z)

The HISTORY of

Guy, *Earl of* Warwick.

CHAP. I.

Guy's Praise. He falls in Love with fair Phillis.

IN the bleſſed Time when Athelſtone wore the Crown of the Engliſh Nation, Sir Guy (Warwick's Mirror and all the World's Wonder) was the chief Hero of the Age; whoſe Proceſs ſo ſurpaſſed all his Predeceſſors, that the Trump of Fame ſo loudly ſounded Warwick's Praiſe, that Jews, Turks, and Infidels became acquainted with his Name.

But

```
<text xml:lang="en">
  <front>
<!-- ... -->
  </front>
  <body>
    <div type="chapter" xml:id="G1711_01">
      <head>CHAP. I.</head>
      <head>Guy's Praise. He falls in
Love with fair Phillis.</head>
      <p>IN the blessed Time when
Athelstone wore the Crown of the
English Nation ... </p>
<!-- ... -->
    </div>
    <div type="chapter" xml:id="G1711_02">
      <head>CHAP. II.</head>
      <head>Guy courts fair Phillis,
she at first denies, but afterwards
grants his
          Suit, on Conditions which
he accepts. </head>
      <p>GUY immediately advanced to
fair Phillis, who was reposing
herself in an Arbour
          ... </p>
<!-- ... -->
    </div>
  </body>
```

Components of a division

What do divisions contain (apart from other divisions) ?

- One or more headings tagged with <head> at the start or <trailer> at the bottom
- Prose, which may be organized as a sequence of paragraphs (<p>) or lists (<list>) containing <item>s
- Poetry, divided into lines (<l>), optionally grouped into stanzas of various kinds (<lg>)
- Drama, divided into speeches (<sp>), containing an optional speaker label, followed by a mix of <p> or <l> elements, optionally mixed up with stage directions (<stage>)

Poetic example

```
<lg type="sonnet">
  <head>130</head>
  <l>My Mistres eyes are nothing like the Sunne,</l>
  <l>Currall is farre more red, then her lips red,</l>
  <l>If snow be white why then her brests are dun:</l>
  <l>If haires be wiers, black wiers grow on her head:</l>
  <l>I have seene Roses damaskt, red and white,</l>
  <l>But no such Roses see I in her cheekes,</l>
  <l>And in some perfumes is there more delight,</l>
  <l>Then in the breath that from my Mistres reekes.</l>
  <l>I loue to heare her speake, yet well I know.</l>
  <l>That Musicke hath a farre more pleasing sound:</l>
  <l>I graunt I never saw a goddesse goe,</l>
  <l>My Mistress when she walkes treads on the ground.</l>
  <lg type="couplet">
    <l>And yet by heaven I thinke my love as rare,</l>
    <l>As any she beli'd with false compare.</l>
  </lg>
</lg>
```

Dramatic example

```
<div type="scene">
<!-- ... -->
 <sp>
  <speaker>Vladimir</speaker>
  <p>Pull on your trousers.</p>
 </sp>
 <sp>
  <speaker>Estragon</speaker>
  <p>You want me to pull off my trousers?</p>
 </sp>
 <sp>
  <speaker>Vladimir</speaker>
  <p>Pull <emph>on</emph> your trousers.</p>
 </sp>
 <sp>
  <speaker>Vladimir</speaker>
  <p>
   <stage>(realizing his trousers are down)</stage>.
True</p>
 </sp>
 <stage>He pulls up his trousers</stage>
 <sp>
  <speaker>Vladimir</speaker>
  <p>Well? Shall we go?</p>
 </sp>
 <sp>
```

Another dramatic example

```
<sp>
  <speaker>Alceste.</speaker>
  <l>Non : j' ai résolu de n' en pas faire un pas.</l>
  <l part="I">J' ai tort, ou j' ai raison.</l>
</sp>
<sp>
  <speaker>Philinte.</speaker>
  <l part="F">Ne vous y fiez pas.</l>
</sp>
<sp>
  <speaker>Alceste.</speaker>
  <l part="I">Je ne remuerai point.</l>
</sp>
<sp>
  <speaker>Philinte.</speaker>
  <l part="F">Votre partie est forte,</l>
  <l part="I">et peut, par sa cabale, entraîner...</l>
</sp>
<sp>
  <speaker>Alceste.</speaker>
  <l part="F">Il n' importe.</l>
</sp>
<sp>
  <speaker>Philinte.</speaker>
  <l part="I">Vous vous tromperez.</l>
</sp>
<sp>
  <speaker>Alceste.</speaker>
  <l part="F">Soit. J' en veux voir le succès.</l>
</sp>
```

Typical textual components

- Strings of words which are typographically distinct (by means of font change, size, capitalization, etc.) for various reasons
- “data-like” or "non-lexical" words we may wish to single out for special processing (names of people or places, dates, numbers ...)
- editorial interventions (corrections, normalisations, additions or deletions in the source ...)
- links and cross references
- lists, notes, annotations, pictures, tables, bibliographic citations...
- paratextual features, such as running heads, catchwords, page numbers, etc.
- interpretive features, such as metaphors, linguistic or stylistic analyses, quotations, allusions etc.

EI provides tags to describe all these and many others...

Rendition and its interpretation

Part of a text is set in a different font : do we care ?

Mais qu'est-ce que cette politesse ? Comment être poli ? Et avec qui ? Il nous le dit dans *Histoires*. Un marchand de pierres à briquet « *élève la voix le soir dans les couloirs de la station Javel* ». Le poème ajoute : « *Et ses grands écarts de langage Déplaisent à la plupart des gens* ». Bien sûr, mais Prévert pense que la *parole* convainc, et il continue « *mais la brûlure de son regard Les rappelle à de bons sentiments* », plus prompts à l'accord que les gens du *Dîner de Têtes*, « *ceux qui mettent un loup sur leur visage quand ils mangent du mouton* », les « *cartonnés officiels* », mais *le Dîner de Têtes* est un des premiers écrits de Prévert.

Rendition and its interpretation - 1

Several policies are possible :

- indicate the presence of a typographic change by using the <hi> (highlighted) element

```
<p>... mais Prévert pense que la <hi>parole</hi> convainc, et il continue «
<hi>mais la brûlure de son regard Les rappelle à de bons sentiments</hi> », plus
prompts à l'accord que les gens du <hi>Dîner de Têtes</hi>, ....</p>
```

- interpret the motivation for the typographic change and indicate that : in this case, there is an <emph> (emphasis), a <quote> (citation) and a <title> (titre):

```
<p>... mais Prévert pense que la <emph>parole</emph> convainc, et il continue «
<quote>mais la brûlure de son regard Les rappelle à de bons sentiments</quote>
», plus prompts à l'accord que les gens du <title>Dîner de Têtes</title>,
....</p>
```

Rendition and its interpretation - 2

Such interpretation may not always be easy or non-controversial

And here note, for a Caution againſt *Extravagance*, and for encouragement to *Frugality* and good Husbandry in all People, eſpecially *Youth*,

That every Penny any Perſon ſpends idly, would purchaſe a Yard (that is three foot) ſquare, and ſomewhat above, of as good Land as moſt in England, to him and his Heirs for ever.

Rendition and its interpretation - 3

- It is therefore useful to combine both levels ...

```
<p>... mais Prévert pense que la <emph rend="italic">parole</emph> convainc,
et il continue « <q rend="italic">mais la
    brûlure de son regard Les rappelle à de bons sentiments</q> », plus prompts
à
l'accord que les gens du <title rend="italic">Dîner de Têtes</title>, ....</p>
```

- The values used for @rend are not defined by the TEI but by each project
- Other similar attributes @style and @rendition may be used in preference to define rendition using a externally defined standard such as CSS.

Phrases in foreign languages

- The @xml:lang attribute can be attached to any TEI element
- The <foreign> element can be used if no other more meaningful tag is appropriate
- The language is identified using an ISO 639 language code
- Language identification includes writing system and dialect variation

```
<q xml:lang="fr">As-tu lu <title xml:lang="de">Die
Dreigroschenoper </title>? </q>
```

Names of people and places

The <rs> (referring string) can be used for any generic reference :

```
<p> "My dear <rs>Mr. Bennet</rs>,"
said <rs>his lady</rs> to him one day, "have you heard that <rs>Netherfield
Park</rs>
is let at last?"</p>
```

The elements <name>, <persName>, or <placeName> are more specific :

```
<p> "My dear <persName>Mr.
    Bennet</persName>," said <rs>his lady</rs> to him one day, "have you heard
that
<placeName>Netherfield Park</placeName> is let at last?"</p>
```

We (Magdalena) will discuss this in (much) more detail tomorrow!

Encoding the physical structure

- Usually, the textual structure (chapters, paragraphs) is considered more important than the physical structure (gatherings, pages, lines).
- However, it is also common to mark the start of gatherings, pages, columns, lines, etc. since they are useful as reference points
- 'milestone' elements such as <gb>, <pb>, <cb> and <lb> are used for this purpose
- Other methods are possible with <sourceDoc> and <facsimile>

```
<p>This paragraph begins on page 11 and continues
<pb  n="12"  facs="http://myPages.org/page12.png"/> on page 12 where it
finishes.</p>
<!-- more text on page 12 here -->
```

Quotation

The TEI distinguishes a variety of 'distancing' elements, many of which will usually be realised with quote marks or italics :

- <q> (anything enclosed in quotation marks for any reason)
- <said> (speech or thought)
- <quote> (passage attributed to an external source)
- <soCalled> (pseudo quotation) and <mentioned> (word being discussed rather than used)
- <cit> (groups a quotation and its source)

```
<quote>
  <l>...  How Earth herself empowered
     him with her trick,</l>
  <l>Gave him the grip and stringency of Winter,</l>
  <l>And all
     the ardour of th' invincible Spring;</l>
  <bibl>
    <author>Wilfred
         Owen</author>
    <title>Letter to Leslie Gunston / The Wrestler</title>
    <date  when="1917-07"/>
  </bibl>
</quote>
```

Temporal expressions

Dates and times can appear in many forms, and may use many calendar systems. TEI provides attributes to normalise all such expressions, by default using W3C conventions.

```
<date when="1917-07">July 1917.<lb/> Wednesday</date>
```

```
<date when="1980-02-21" xml:lang="fr">21
février 1980</date>
```

```
<date when="2001-09-11T12:48:00">0948 EST
on 11 September 2001</date>
```

```
 Décret de la Convention <date when="1794-10-30">9 Brumaire
An III</date>
```

Imprecise dates and date ranges can also be expressed.

Simple Linking

- <ptr>: (defines a pointer to another location)
- <ref>: (defines a reference to another location, with optional linking text)
- Both elements have:
  - @target attribute taking a URI reference
  - @cRef attribute for canonical referencing schemes

```
 See <ref target="#Section12">section 12
on page 34</ref>.
```

```
 See <ptr target="#Section12"/>.
```

```
 The <ref target="http://www.bbc.co.uk/">BBC web site</ref>
has a good sports section
```

Graphics

- <graphic> (indicates the location of an inline graphic, illustration, or figure)
- <binaryObject> (encoded binary data embedding a graphic or other object)
- The figures module provides <figure> and <figDesc> for more complex graphics
- The @facs attribute links to an image of the source being encoded
- The @url attribute (on <graphic>) links to an image forming part of the source

For example

```
<pb n="05" facs="page_05.png"/>
<p>Phillis interrupted him,
saying, Cease bold Youth,
leave off this passionate
Address :—You are but young
and meanly born, and unfit for
my Degree; I would not my
Father should know this
Passion.
</p>
<figure>
 <graphic url="page_05_detail.png"
 <figDesc/>
</figure>
<p> Guy, thus discomfited,
lived like one distracted,
wringing his Hands, resolving
to travel through the World to
gain the Love of Phillis, or
end his Days in Misery. Long
may Dame Fortune frown, but
n her Course is run she
ds a Smile to
>
```

(5)

Phillis interrupted him, ſaying, Ceaſe bold Youth, leave off this paſſionate Addreſs :—You are but young and meanly born, and unfit for my Degree; I would not my Father ſhould know this Paſſion.

Guy, thus diſcomfited, lived like one diſtracted, wringing his Hands, reſolving to travel through the World to gain the Love of Phillis, or end his Days in Miſery. Long may Dame Fortune frown, but when her Courſe is run ſhe ſends a Smile to

Bibliographies

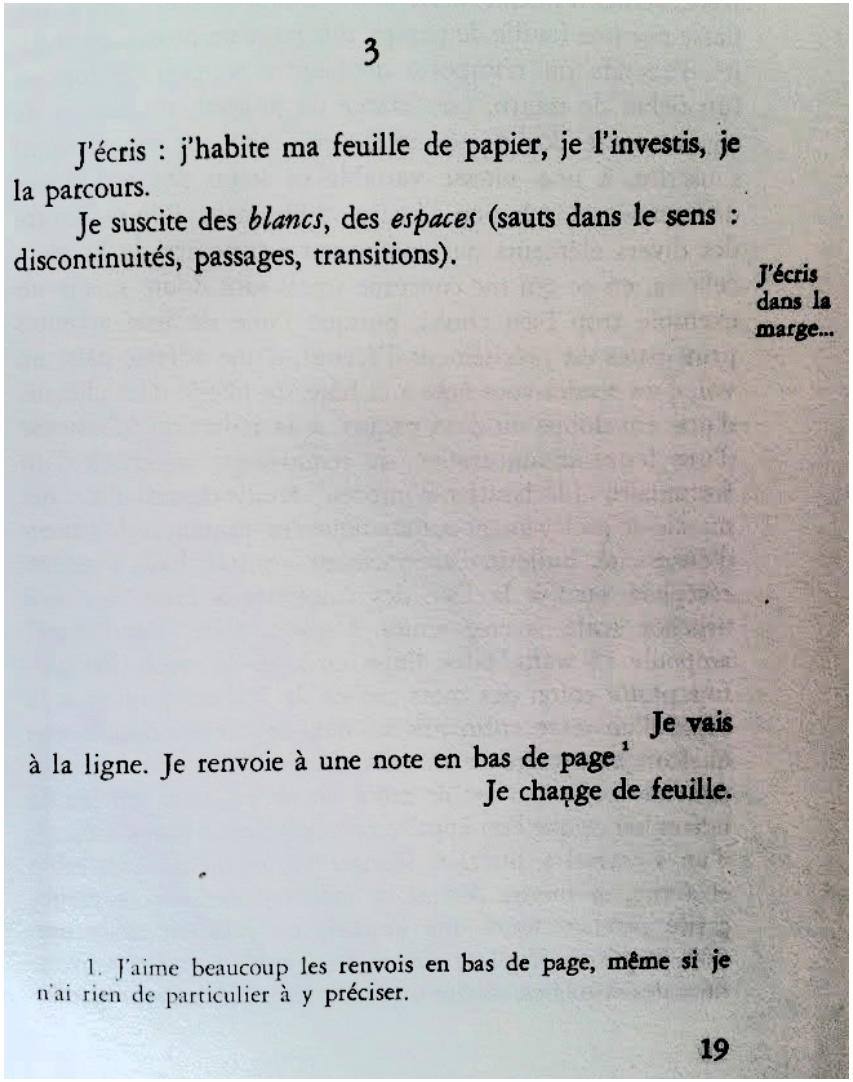
- A bibliographic reference (<bibl>) may be structured or unstructured
- Its components, if distinguished, may use specialised elements such as <author>, <editor>, <title>, <pubPlace>, <publisher>, <series> etc.
- There is no requirement to respect presentational conventions (punctuation, order, layout etc) within a <bibl> since these vary so much, though it can be done where the intention is to represent accurately the format of the source

For example

```
<bibl xml:id="MK_73">Sturm, U. & Drang, F. :
<title>Musikalische Katastrophe</title>.
(Berlin, W. de Gruyter, 1973)</bibl>
```

```
<bibl xml:id="MK73">
 <author>Sturm, U.</author>
 <author>Drang, F.</author>
 <title xml:lang="de" level="m">Musikalische Katastrophe
</title>
 <pubPlace>Berlin</pubPlace>
 <publisher>W. de Gruyter</publisher>
 <date>1973</date>
</bibl>
```

A (French) word on foot notes



3

J'écris : j'habite ma feuille de papier, je l'investis, je la parcours.

Je suscite des *blancs*, des *espaces* (sauts dans le sens : discontinuités, passages, transitions).

J'écris dans la marge...

Je vais à la ligne. Je renvoie à une note en bas de page[1]

Je change de feuille.

1. J'aime beaucoup les renvois en bas de page, même si je n'ai rien de particulier à y préciser.

19

```
<div n="3">
  <p>J'écris : j'habite ma feuille de
papier, je l'investis, je la
parcours.</p>
  <p>Je suscite des <hi>blancs</hi>, des
<hi>espaces</hi> (sauts dans le sens :
    discontinuités, passages,
transitions).
<note place="margin">J'écris dans la
       marge...</note>
  </p>
  <p rend="indent3">Je vais <lb/>à la
ligne. Je renvoie à une note en bas de
page<note place="foot">J'aime beaucoup
les renvois en bas de page, même si je
n'ai rien de
       particulier à y préciser.</note>
  </p>
  <p rend="indent2">Je change de
feuille.</p>
</div>
<pb n="20"/>
```

Do you feel a little overwhelmed ?

All this is just one way of using the TEI.

- TEI is a modular system. You use it to create an encoding system that reflects your own needs, by choosing from the TEI's pre-defined modules
- Each module defines a set of elements and their attributes
- you can choose just the elements you want, and also (within limits) change their properties
- you can add in non-TEI elements, either from other standards or completely new

Define your goals clearly before trying to use the TEI !