

03 – Introduction aux richesses de la TEI

Par exemple

A MONSEI-

GNEVR LE REVE-
rendissime Cardinal
du Bellay.

S.



EV le Personnage,
que tu ioues au Specta-
cle de toute l'Europe,
uoyre de tout le Mon-
de en ce grand Thea-
tre Romain, uen tant
d'affaires, & telz, que
seul quasi tu soutiens: ô
l'Honneur du sacré Col-

lege! pecheroy'-ie pas (comme dit le Pindare
Latin) contre le bien publicq', si par longues
paroles i' empeschoy' le tens, que tu donnes au

Qu'est ce qu'on balisera?

Balisage du mise en evidence

A MONSEI-
GNEVR LE REVE-
rendissime Cardinal
du Bellay.
S.



EV le Personnage,
que tu ioues au Specta-
cle de toute l'Europe,
uoire de tout le Mon-

```
<pb n="4"/>A MONSEI-  
<lb/>GNEVR LE REVE-  
<lb/>rendissime Cardinal  
<lb/>du Bellay.  
<lb/>S
```

```
<lb/>  
<c rend="lettrine">V</c>EV le  
Personnage,  
<lb/>que tu ioues au Specta-  
<lb/>cle de toute l'Europe...
```

Balisage de structure

A MONSEI-
GNEVR LE REVE-
rendissime Cardinal
du Bellay.
S.

 EV le Personnage,
que tu ioues au Specta-
cle de toute l'Europe,
noyre de tout le Mon-

```
<div type="dedicace">  
  <head>A MONSEIGNEUR LE  
  REVERENDISSIME CARDINAL DU  
  BELLAY</head>  
  <salute>S<ex>ire</ex>  
  </salute>  
  <p>  
    <c rend="lettrine">V</c>EU  
    le Personnage, que tu jous  
    au Spectacle de toute  
    l'Europe... </p>...  
</div>
```

Mais on peut aller plus loin...

```
<pb n="4"/>
<s>
  <w pos="PPJ" lemma="voir">VEU</w>
  <w pos="ART" lemma="le">le</w>
  <w pos="SBC" lemma="personnage">Personnaige</w>
  <pc>,</pc>
  <w pos="C00" lemma="que">que</w> ...
</s>
```

ou bien

```
<s>
  <choice>
    <reg>Vu</reg>
    <orig>Veu</orig>
  </choice> le
  <choice>
    <reg>Personnage</reg>
    <orig>Personnaige</orig>
  </choice>, que tu joues
  au Spectacle...
</s>
```

à ne rien dire de ...

```
<head> A MONSEIGNEUR LE REVERENDISSIME  
<persName ref="#dubellay03">CARDINAL DU  
  BELLAY</persName>  
</head>  
<!-- .... -->  
<person xml:id="dubellay03">  
  <persName>Jean du Bellay</persName>  
  <birth>  
    <date>1492</date>  
    <placeName>Souday</placeName>  
  </birth>  
  <death>  
    <date when="15600216">16 February 1560</date>  
    <placeName>Roma</placeName>  
  </death>  
<!-- ... -->  
</person>
```

Les Cinq Étapes du chemin de l'Éveil TEI

- ① Modélisation : quels sont les buts et les objets de votre projet ?
- ② Orientation : est-ce que vos objets sont déjà reconnus par la TEI ?
- ③ Déclaration : comment créer une spécification TEI-conforme ?
- ④ Documentation : comment documenter vos pratiques TEI ?
- ⑤ Validation : comment valider vos efforts ?

Modélisation

La modélisation de vos données est une préalable essentiel. Que vous vous serviez de UML, de RDBMS, de SKOS, ou de quoique ça soit d'autre, si vous n'avez pas un modèle explicite des choses que vous espérez gérer, vous aurez de grands difficultés.



Orientation



Comment s'orienter ?

Les *TEI Guidelines*, imprimées ou en ligne, se divisent en deux grandes parties :

- une séquence de 23 chapitres, dont la plupart correspondent à un module
- des listes alphabétiques de définitions exhaustives de classes (183), d'éléments (546), d'attributs (470), de macros(8), et de types de données (28).

Comment savoir quel élément (etc.) choisir pour telle ou telle entité identifiée dans votre analyse préalable ? Comment savoir que vous avez besoin d'une 'licorne' ?

La triste vérité...

- Désolé, mais il n'y a aucune méthode scientifique ; aucun raccourci pour cela...
- Il faut étudier les exemplaires et les définitions pour savoir si cet élément TEI si prometteur s'applique en effet à votre cas
- comme c'est le cas pour d'autres types de mapping...

Pour chacun(e) des entités/concepts identifiés dans votre modèle, il faut décider :

- quel objet TEI lui correspond parfaitement ? (s'il existe)
- quel objet TEI lui ressemble et quelles petites modifications seraient nécessaires pour qu'il lui corresponde ?
- quelle lacune TEI votre analyse vous permettra de corriger (oui, elles existent les licornes !)

Structuration

- C'est un livret de la bibliothèque bleue – un 'chapbook' – organisé de manière traditionnelle, avec des pièces liminaires (une page de titre, une table des matières) et plusieurs chapitres
- chacun des chapitres contient :
 - un titre et un sommaire
 - des paragraphes narratifs
 - des passages de dialogue
 - des noms de lieux, noms de personnes, noms d'événements, etc.
- chaque page est dotée de matières paratextuelles (signatures, etc.)
- Et plusieurs images

Structure(s) d'un document XML-TEI

- Chaque document TEI est représenté par un élément `<TEI>`, regroupant le contenu d'un document avec un entête de métadonnées (`<teiHeader>`) qui le décrivent
- On peut également regrouper des `<TEI>` dans un `<teiCorpus>`

Le terme *document* n'a aucune définition formelle, mais d'habitude il correspond à un objet complet (un livre par ex). D'autres possibilités ne sont pas exclus...

Un document TEI peut contenir...

- un ensemble de représentations visuelles (<facsimile>)
- ou une transcription textuelle (<text>)
- ou une transcription documentaire (<sourceDoc>)
- ou ces trois types d'objet conjointement

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- obligatoire : métadonnées -->
  </teiHeader>
  <facsimile>
    <!-- facultatif : rassemble des images de l'objet-->
  </facsimile>
  <sourceDoc>
    <!-- facultatif : transcription documentaire de l'objet -->
  </sourceDoc>
  <text>
    <!-- facultatif : transcription textuelle de l'objet -->
  </text>
</TEI>
```

Un TEI Corpus par contre...

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- obligatoire -->
  </teiHeader>
  <TEI>
    <!-- obligatoire -->
  </TEI>
  <!-- encore d'autres éléments <TEI> -->
</teiCorpus>
```

Cette structure servirait par exemple à une collection de livres de ce genre

L'entête minimale

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>titre de la ressource</title>
    </titleStmt>
    <publicationStmt>
      <p>agence de distribution de la source</p>
    </publicationStmt>
    <sourceDesc>
      <p>provenance de la source</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

L'élément <text>

Un texte peut être simple ou complexe :

- un <text> simple est un objet autonome, en lui-même complet, par ex un roman, un poème
- un <text> complexe ou composé, rassemble d'autres textes, par exemple une anthologie, une collection de poésies

TEI text structure (1)

Un document simple :

```
<text>
  <front>
<!-- facultatif -->
    </front>
    <body>
<!-- obligatoire -->
    </body>
    <back>
<!-- facultatif -->
    </back>
</text>
```

TEI text structure (2)

Un document composé :

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

Entrons dans le <body> d'un <text>...

D'habitude, on y trouve des *divisions*

- qui s'emboîtent
- l'attribut *@type* sert à distinguer les divisions par fonction ou par niveau
- l'attribut *@n* fournit un nom ou un nombre pour une division déjà numérotée d'une manière quelconque
- l'attribut *@xml:id* fournit une identifiant unique
- l'attribut *@xml:lang* identifie la langue (humaine) du contenu de la division

Ces deux derniers sont deux des attributs dits *globaux* qui peuvent s'attacher à n'importe quel élément TEI

Composants des divisions

Nous reconnaissons facilement...

- éléments liminaux spécialisés : `<head>`, `<trailer>`, etc.
- en prose : les `<p>`, `<list>` etc.
- en poésie : les `<lg>`, `<l>` ...
- en drame : les `<sp>`, `<stage>` etc.
- partout : les `<note>`, les `<bibl>`, les `<figure>` ...

Composants des composants des divisions

On peut également baliser...

- éléments de repérage : les `<lb>`, `<pb>` etc.
- éléments et attributs de mise-en-forme : `<hi>`, `@rend`, `@rendition` etc.
- éléments sémantiques : `<date>`, `<name>`, `<placeName>` etc.
- éléments para-textuels : `<sic>`, `<corr>`, `<add>` etc.
- etc., etc., etc.

Notre exemple (1)

```
<text>
  <front>
    <titlePage>
      <docTitle>cy commence Guy de Wa- ruich
chevalier Dagleterre qui en son temps fit
      plusieurs prouesses et con-
questes en Allemaigne/ytalie et
Da<damage/>emarche. Et
      aussi sur les infidel- les ennemys
de la chrestiente Comme pourrez veoir plus a
      plain en ce pre- sent livre
Imprime nouvellement a Paris </docTitle>
      <figure>
        <graphic url="gdw-tp-1"/>
        <head>FRANCOYS REGNAULT</head>
      </figure>
      <imprint>cum privilegio</imprint>
    </titlePage>
  <!-- ... -->
</front>
</text>
```



Notre exemple (2)

```
<text>
<front>
<!-- ... -->
  <div type="prolog" xml:id="p1">
    <head>Cy commence le prologue de
lacteur</head>
    <figure>
      <graphic url="gdw-p01.png"/>
    </figure>
    <p>
      <g type="lettrine" ref="#ornA">A</g>u
temps du roy A- thelstā prince de no- ble
memoire regnat en souverainete ou royaulme
Dangle- terre <abbr>ape</abbr>lan de lin-
carnation notre sei gne<am>r</am> Jesuchrist
qua tre cens ans et vingtquatre ... </p>
    <trailer>Ici fine le prologue</trailer>
  </div>
</front>
<!-- ... -->
</text>
```



Notre exemple (3)

```
<text>
<!-- ... -->
<body>
  <fw>feuillet premier.</fw>
  <div type="chap" xml:id="C1">
    <head>Comment le conte Roalt avoit une
belle fille nommee Felisse sur- montant en
beautee toutes autres dames et damoiselles et
envoya querir Guyonet filz de son seneschal
pour demeurer en sa court.</head>
    <figure>
      <graphic url="gdw-p02.png"/>
    </figure>
    <p>
      <g type="lettrine" ref="#ornE">E</g>n
icelle honorable sai- son a regne dudit roy
A- thelstā estoit ou royaul- me dangleterre
ung tres puissant cont nomme Roalt
<abbr>leql</abbr> avoit la sei- gneurie de la
conte de waruich ...</p>
    </div>
  </body>
</text>
```



Composants des divisions

Qu'est-ce qu'on trouve dans une division ?

- Un ou plusieurs titres (au début et/ou à la fin), balisés avec `<head>` ou `<trailer>`
- de la prose, divisée en paragraphes (`<p>`) ou listes (`<list>`) contenant des `<item>` et facultativement des `<label>`
- de la poésie, divisée en vers (`<l>`), facultativement regroupés en strophes (`<lg>`)
- des dialogues divisés en énoncés (`<sp>`), contenant une mélange de `<p>` ou de `<l>` facultativement entrecoupée de didascalies (`<stage>`)

... et bien sûr des sous-divisions

Exemple prosaïque...

```
<div n="1">
  <head>La Fête</head>
  <p>C'était à Mégara, faubourg de Carthage, dans les
    jardins d'Hamilcar.</p>
  <p>Les soldats qu'il avait commandés en Sicile se
    donnaient un grand festin pour célébrer le jour
    anniversaire de la bataille d'Eryx, et comme le
    maître était absent et qu'ils se trouvaient nombreux,
    ils mangeaient et ils buvaient en pleine liberté.</p>
  <!-- ... -->
</div>
```

Exemple poétique

```
<div type="sonnet">
  <lg type="quatrain">
    <l>Les amoureux fervents et les savants austères</l>
    <l>Aiment également, dans leur mûre saison,</l>
    <l>Les chats puissants et doux, orgueil de la maison,</l>
    <l>Qui comme eux sont frileux et comme eux sédentaires.</l>
  </lg>
  <lg type="quatrain">
    <l>Amis de la science et de la volupté</l>
    <l>Ils cherchent le silence et l'horreur des ténèbres ;</l>
    <l>L'Èrèbe les eût pris pour ses coursiers funèbres,</l>
    <l>S'ils pouvaient au servage incliner leur fierté.</l>
  </lg>
  <lg type="tercet">
    <l>Ils prennent en songeant les nobles attitudes</l>
    <l>Des grands sphinx allongés au fond des solitudes,</l>
    <l>Qui semblent s'endormir dans un rêve sans fin ;</l>
  </lg>
  <lg type="tercet">
    <l>Leurs reins féconds sont pleins d'étincelles magiques,</l>
    <l>Et des parcelles d'or, ainsi qu'un sable fin,</l>
    <l>Étoilent vaguement leurs prunelles mystiques.</l>
  </lg>
</div>
```

Exemples dramatiques (1)

```
<sp who="#V">
  <speaker>Vladimir</speaker>
  <p>Relève ton pantalon.</p>
</sp>
<sp who="#E">
  <speaker>Estragon</speaker>
  <p>Que j'enlève mon pantalon ?</p>
</sp>
<sp who="#V">
  <speaker>Vladimir</speaker>
  <p>
    <emph>RE</emph>lève ton pantalon.</p>
</sp>
<sp who="#E">
  <speaker>Estragon</speaker>
  <p>C'est vrai. </p>
  <stage>Il relève son pantalon. Silence.</stage>
</sp>
<sp who="#V">
  <speaker>Vladimir</speaker>
  <p>Alors, on y va ?</p>
</sp>
<sp who="#E">
  <speaker>Estragon</speaker>
  <p>Allons-y.</p>
</sp>
<stage>Ils ne bougent pas.</stage>
<list type="speakers">
  <item xml:id="V">Vladimir</item>
  <item xml:id="E">Estragon</item>
</list>
```

Exemples dramatiques (2)

```
<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 suite.</l>
</sp>
```

Composants textuels typiques

- 1 des séquences de mots qui sont (par convention) mises en évidence, par exemple pour leur donner une emphase, parce qu'ils sont en langue étranger, ou cités etc,
- 2 des mots qui pourraient être d'intérêt non-lexicale "data-like" (des noms de lieu ou de personne, des dates, des chiffres...)
- 3 des indications d'interventions éditoriales (corrections, normalisations, ajouts, ratures...)
- 4 des références et des liens
- 5 listes, notes, annotations, graphies, tables, citations bibliographiques...
- 6 indications de structuration dans une autre hiérarchie, par exemple frontières de pages

TEI fournit des balises pour toutes ces choses : mais il faut faire un choix, selon les objectifs de son projet

(1) La mise en évidence et son interprétation

On voit une partie en italique ou en petites capitales : et on se demande pourquoi... par exemple :

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.

Mise en évidence, & son interprétation (1)

On signale la présence d'une mise en évidence sans distinction particulière avec `<hi>` (highlighted : distinction typographique)

```
<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>
```

on essaie de distinguer l' **objectif sémantique** de ces mises en évidence : dans le cas présent, il y a un `<emph>` (emphasis linguistique), une `<quote>` (citation) et un `<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>
```

La mise en évidence et son interprétation - 2

on peut combiner ces deux niveaux de description...

```
<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>
```

Les valeurs possibles de l'attribut *@rend* sont à définir pour chaque projet.

Des attributs complémentaires *@style* ou *@rendition* permettent l'usage des systèmes déjà standardisés, notamment le CSS.

Phrases en langues étrangères

- L'attribut `@xml:lang` peut s'attacher à tout élément TEI
- L'élément `<foreign>` est disponible s'il n'y a pas d'autre sémantique à signaler
- La langue est identifiée par son code normalisé ISO

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

Tout d'un coup il exclame `<foreign xml:lang="de">` himmel ! `</foreign>`

(2) Noms de lieux et de personnes

La balise `<rs>` (referring string) est très générique :

```
<p>Le premier dîner que <rs type="person">M. de Norpois</rs>  
fit à la maison, une année où je jouais  
encore aux <rs type="place">Champs-Élysées</rs>,  
est resté dans ma mémoire, parce que l'après-midi  
de ce même jour fut celui où j'allai enfin entendre  
<rs type="person">la Berma</rs>, en « matinée »,  
dans <title>Phèdre</title>...</p>
```

Les balises `<name>`, `<persName>`, ou `<placeName>` sont plus spécifique :

```
<p>Le premier dîner que  
<persName ref="http://fr.wikipedia.org/wiki/Marquis_de_Norpois">M. de  
Norpois</persName> fit à la maison, une année où je jouais  
encore aux <placeName>Champs-Élysées</placeName>, est resté  
dans ma mémoire, parce que l'après-midi de ce même jour fut  
celui où j'allai enfin entendre <name>la Berma</name>, en  
« matinée », dans <title>Phèdre</title>...</p>
```

Nous allons traiter ce sujet plus tard plus en détails.

Expressions temporelles

Les dates qui figurent dans les textes manifestent beaucoup de variations lexicales, pour ne rien dire des variations des systèmes de calendrier. Pour comparer et contrôler de telles expressions d'une manière efficace, il faut les normaliser. La TEI propose un élément `<date>` pour identifier ces expressions, et plusieurs attributs pour leur normalisation.

```
<date when="1980-02-21">21 février
```

```
1980</date>
```

```
<date when="1990-02-21" xml:lang="en">21 September,
```

```
1990</date>
```

```
<date when="2001-09-11T12:48">11
```

```
septembre, neuf heures moins douze, EST</date>
```

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

(3) Les interventions éditoriales

En transcrivant une source primaire on pourrait vouloir

- respecter, normaliser, ou corriger l'orthographe originel
- signaler la présence des ajouts, ratures, etc.
- développer ou signaler les abréviations
- indiquer des leçons varies pour une source difficile a lire ou existant en plusieurs versions
- indiquer des changements de main ou d'écriture

Nous allons traiter plus tard ce sujet avec plus de détails.

(4) Les références croisées et les liens

Pour relier deux endroits d'un ou plusieurs textes, on utilise l'un des éléments `<ptr>` ou `<ref>`.

Voir `<ref target="#pv1894_61">` procès-verbal
no 61`</ref>`.

Voir `<ptr target="#pv1894_61"/>`.

L'attribut `@target` spécifie l'objet ciblé par la référence, que ce soit en fournissant l'identifiant d'un autre élément dans le même document, ou en fournissant une adresse complète sur le Web (un URL)

Voir `<ptr target="http://www.example.org/example.xml#pv1894_61"/>`

Dans ces exemples, on pointe sur la section suivante :

```
<div xml:id="pv1894_61">  
<!-- ... -->  
</div>
```

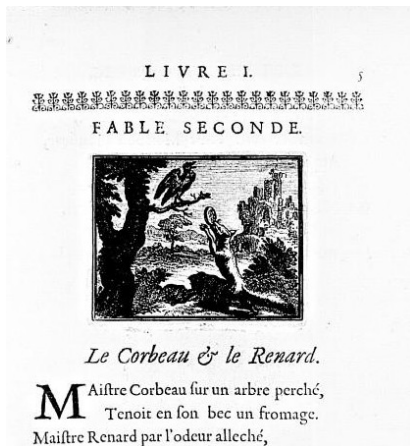
Plusieurs autres attributs de type 'pointeur' sont disponibles, fournissant une gamme de type de liens plus spécialisés.

(5) Les images

- `<graphic>` précise la location d'une image, illustration, etc. faisant partie du texte)
- `<figure>` précise une partie iconographique plus complexe, rassemblant par exemple plusieurs images, un titre, des `<figDesc>` pour les décrire
- L'attribut `@facs` fournit un lien externe sur une représentation supplémentaire de la source encodée
- L'attribut `@url` par contre (sur `<graphic>`) fournit un lien sur une image qui est intégrale à cette source

Par exemple

```
<pb n="5" facs="page_A3r.png"/>
<fw>5</fw>
<div>
  <head>LIVRE I. FABLE
  SECONDE.</head>
  <figure>
    <graphic url="I2dtl.png"/>
    <figDesc>Un renard regarde
    un arbre dans lequel se trouve
    un corbeau. Le renard tient un
    fromage dans la bouche. En
    arrière plan, un
    château.</figDesc>
  </figure>
  <head>Le Corbeau & le
  Renard.</head>
  <l>Maistre Corbeau sur un
  arbre perché...</l>
</div>
```



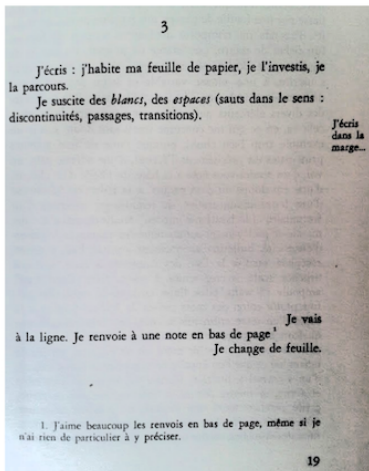
Les citations bibliographiques

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

- Une référence bibliographique peut être plus ou moins structurée, en se servant d'éléments tels que `<author>`, `<editor>`, `<title>`, `<pubPlace>`, `<publisher>`, `<series>` etc.
- les conventions de présentation des références ne sont pas forcément à encoder

```
<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>
```

Et pour terminer, les notes



```
<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"/>
```

Vous vous sentez un peu accablé ?

Tout ceci n'est qu'une des manières de se servir de la TEI.

- TEI est un système *modulaire*. On s'en sert pour créer un système d'encodage selon ses propres besoins, en sélectionnant des *modules* spécifiés par la TEI
- Chaque module définit un groupe d'éléments (et leurs attributs)
- on peut sélectionner les éléments voulus, et même en modifier des propriétés
- on peut y mélanger des éléments nouveaux, ou bien natifs ou bien d'autres standards

Il faut définir ses objets *avant* de se servir la TEI !

La TEI facilite un balisage 'intelligent'

La TEI s'applique à l'encodage des...

- composants structuraux et fonctionnels d'un texte
- transcriptions diplomatiques des sources historiques, des images, des annotations
- liens, correspondances, alignements
- données et entités : par exemple de temps, personnes, lieux ou événements
- annotations péri-textuelles et méta-textuelles (correction, suppression, ajouts)
- analyses linguistiques
- métadonnées de plusieurs types
- ... et définitions formelles de schéma XML !

On est vraiment gâté par l'étendue des choix possibles...

TEI modules (1)

Vous aurez assurément besoin des éléments de ces modules :

module	contenant	P5	P3	
core	features common to almost every kind of document	éléments communs à presque tout type de documents	82	71
figures	tables, formulas, figures	tables, formules et figures	7	6
header	bibliographic and other metadata	les métadonnées bibliographiques et autres	69	57
tei	datatypes, classes, and macros used by every TEI module	0	0	
textstructure	basic structural elements	33	31	

TEI Modules (2)

Pour les manuscrits vous aurez probablement besoin de ceux ci:

module	contents	P5	P3
gaiji	handling special (non-Unicode) characters and glyphs	11	15
msdescription	manuscript and incunabula cataloguing and description	69	0
textcrit	traditional critical apparatus	14	12
transcr	diplomatic or genetic transcription of primary sources	29	10

TEI Modules (3)

D'autres types de document sont envisageables...

module	contents	P5	P3
dictionaries	print dictionaries	35	37
drama	performance texts	17	16
spoken	oral transcription	14	7
tagdocs	markup documentation (ODD)	54	31
verse	poetic texts	4	6

TEI Modules (4)

Si vous faites des analyses spécialisés vous aurez peut être besoin de ceux-ci :

module	contents	P5	P3
analysis	simple analytic mechanisms	11	10
certainty	uncertainty and relevance	3	2
corpus	language corpus metadata	14	26
iso-fs	feature structure based analysis	28	32
linking	links, segments, alignment	11	12
namesdates	names, dates, and named entities	52	30
nets	graphs, networks, and trees	12	12

Niveaux de validation

Un document TEI-XML doit :

- ① respecter les règles syntaxiques d'XML ;
- ② être valide par rapport à un schéma quelconque;
- ③ respecter la sémantique définie de chaque élément TEI utilisé

Un schéma exprime d'une manière formelle qui permet la validation logicielle, une partie importante de ces règles

Un TEI ODD nous permet de construire un schéma adéquat à nos besoins

Comment effectuer notre choix de balises ?

Nous aurons besoins de plusieurs choses :

- un schéma formel (en langue informatique tel que DTD, RELAXNG, W3C Schema, Schematron) qui peut contrôler :
 - quelles balises sont disponibles ?
 - dans quels contextes ?
 - avec quels attributs ?
 - avec quelles valeurs ?
 - en respectant quelles contraintes ?
- une documentation pour expliquer nos principes éditoriaux, nos principes de choix de balises, etc. aux utilisateurs/développeurs :
 - en plusieurs langues naturelles
 - en plusieurs formats bureautiques (PDF, Word, HTML, epub...)
- des outils informatiques pour transformer et gérer nos données XML

Propositions de la TEI

Ayant elle-même ces mêmes besoins, la TEI vous propose :

ODD Un vocabulaire XML pour définir les vocabulaires XML

Stylesheets Un ensemble de feuilles de style XSLT très générique pour la conversion des documents XML TEI

Roma

Un logiciel pour traiter des documents exprimés dans ce vocabulaire, pour en faire ressortir

- des schémas RELAXNG, DTD, etc.
- des manuels "mode d'emploi" en HTML, PDF etc.

OxGarage

Un outil web pour effectuer plusieurs transformations documentaires, par ex

- de DOCX en TEI, et l'inverse
- de HTML en TEI, et l'inverse

Nous essayons le mieux possible d'internationaliser la TEI !

- Suite à un effort financé par le ALLC (entre autres), les définitions, les gloses, et quelques exemples sont traduits en plusieurs langues, notamment en français, mais aussi :
 - en espagnol, en portugais, en allemand, en chinois, en koréen, en japonais...
- Exception faite de quelques épreuves, les chapitres des *Guidelines* actuelles restent jusqu'à présent rédigées en anglais.

L'usage de ODD nous permet de sélectionner une documentation française.

L'idée essentielle de ODD

One Document Does it all

Un vocabulaire spécialisé pour la définition

- des schémas
- des types d'élément XML, indépendant des schémas
- des regroupements de tels éléments, publics ou privés
- des patrons (MLE macros)
- des classes (et sous-classes) d'éléments

Pour la définition des références également, qui permettront de réunir dans un schéma

- des objets identifiables (dans la liste ci-dessus)
- des objets appartenant à d'autres espaces de nom

et qui serait intégrable à un système de balisage documentaire classique

Être conforme à la TEI, ça veut dire quoi ?

- **être honnête** : Les éléments XML qui se déclarent comme appartenant à l'espace de nom TEI doivent respecter les définitions TEI de ces éléments
- **être explicite** : Pour valider un document TEI, un ODD est fortement conseillé, parce que cela mettra en évidence toutes les modifications effectuées
- Tout document valide TEI est valide par rapport au schéma TEI-ALL

L'objet de ces règles est de faciliter le *blind interchange* des documents.