

L'envergure de la TEI

Un tour d'horizon de l'univers TEI

The TEI Guidelines

Un tour d'horizon très rapide du contenu des TEI *Guidelines*...

Dans ses 1400 pp imprimées, vous trouverez :

un lexique 521 éléments, regroupés en 146 classes, et 22 modules

des règles d'usage 7185 lignes de règles syntaxiques en Relax NG

des contraintes formalisées 21 types de données, plusieurs règles formalisées en Schematron

des règles / conventions d'utilisation *beaucoup* de prose

plusieurs exemples d'utilisation dont au moins un par élément

La liste des chapitres (1)

À part les pièces liminaires, les *Guidelines* contiennent

- 1. The TEI Infrastructure
- 2. The TEI Header
- 3. Elements Available in All TEI Documents
- 4. Default Text Structure
- 5. Representation of Non-standard Characters and Glyphs
- 6. Verse
- 7. Performance Texts
- 8. Transcriptions of Speech
- 9. Dictionaries
- 10. Manuscript Description
- 11. Representation of Primary Sources
- 12. Critical Apparatus

...

La liste des chapitres (2)

...

- 13. Names, Dates, People, and Places
- 14. Tables, Formulæ, and Graphics
- 15. Language Corpora
- 16. Linking, Segmentation, and Alignment
- 17. Simple Analytic Mechanisms
- 18. Feature Structures
- 19. Graphs, Networks, and Trees
- 20. Non-hierarchical Structures
- 21. Certainty and Responsibility
- 22. Documentation Elements
- 23. Using the TEI

v. A Gentle Introduction to XML

Un texte classique sur l' SGML, mise à jour pour l'XML.

- Résume tout ce qu'il faut savoir sur l'XML pour se servir des Guidelines.
- Incontournable pour les XML novices
- Traite des principes du balisage, sa représentation en XML, et les possibilités de traitement ainsi fournies

1 - 5 : Les essentiels

1. **The TEI Infrastructure** Définition et explication des classes et des attributs globaux : présentation essentielle de l'architecture informatique de la TEI
2. **The TEI Header** Tout ce qu'il faut savoir sur les métadonnées
3. **Éléments Available in All TEI Documents** Une cinquantaine d'éléments à utiliser dans presque tout document ... le noyau du système
4. **Default Text Structure** Structuration (par défaut) des documents TEI en sections, sous-ensembles, groupements etc.
5. **Representation of Non-standard Characters and Glyphs**
Traitement des caractères et des glyphes non-Unicode. Entre autres, cette section propose l'élément `<g>` qui permet de référencer les caractères et les glyphes non-Unicode et l'élément `<charDecl>` qui permet de les définir.

6. Verse

- Le noyau fournit déjà : `<lg>` (line-groups) et `<l>` (line)
- Ce module rajoute : `<rhyme>` pour la rime d'un vers
- `<caesura>` pour indiquer le lieu de césure du vers
- `<metDecl>` et `<metSym>` pour déclarer un système d'analyse métrique et des attributs `@met` et `@real` pour s'en servir

```
<lg type="quatrain" rhyme="aabccb">
  <l> Les étoiles, points d'or, percent les branches n<rhyme label="a">oires</rhyme> ; </l>
  <l> Le flot huileux et lourd décompose ses m<rhyme label="a">oires</rhyme>
  </l>
  <l> Sur l'océan blé<rhyme label="b">mi</rhyme> ; </l>
  <l>Les nuages ont l'air d'oiseaux prenant la f<rhyme label="c">uite</rhyme>, </l>
  <l>Par moments le vent parle, et dit des mots sans s<rhyme label="c">uite</rhyme>
  </l>
  <l> Comme un homme endor<rhyme label="b">mi</rhyme>. </l>
</lg>
```

7. Performance Texts

- Le noyau fournit déjà : `<sp>`, `<stage>`, `<speaker>`
- Ce module rajoute : `<actor>`, `<camera>`, `<caption>`,
`<castGroup>`, `<castItem>`, `<castList>`, `<epilogue>`, `<move>`,
`<performance>`, `<prologue>`, `<role>`, `<roleDesc>`, `<set>`,
`<sound>`, `<tech>`, `<view>`

```
<performance>
  <p>
    <name type="theatre">Théâtre national du Petit-odéon, Paris</name>, <date>du 10
      Décembre 1974 au 12 Janvier 1975</date>
    <castList>
      <castItem>
        <role>Paul Valéry</role>
        <actor>Michel Duchaussoy</actor>
      </castItem>
      <castItem>
        <role>Joseph</role>
        <actor>Gérard Caillaud</actor>
      </castItem>
    <!-- ... -->
    </castList>
  </p>
  <p>Adaptation et mise en scène de <name>Pierre Franck</name>
</p>
</performance>
```


8. Transcriptions of Speech

Ceci est un module spécialisé pour les transcriptions de discours oraux :

- il fournit des éléments supplémentaires pour le Header (`<broadcast>`, `<equipment>`, `<recording>`, `<recordingStmt>`, `<scriptStmt>`)
- des éléments pour la transcription (`<incident>`, `<kinesic>`, `<pause>`, `<shift>`, `<u>`, `<vocal>`, `<writing>`)

```
<vocal iterated="true">
  <desc>sifflements intermittents</desc>
</vocal>
<u who="#spkr1">si tu te déplaçais</u>
<u trans="latching" who="#spkr2">Joe et moi l'aurions mis entre nous</u>
<list type="speakers">
  <item xml:id="spkr1"/>
  <item xml:id="spkr2"/>
</list>
```

9. Dictionaries

Ce module propose des éléments spécialisés pour les différentes composantes d'un dictionnaire, organisées dans une structure stricte (`<entry>`) ou quasi-aléatoire (`<entryFree>`)

Composants du module : `<case>`, `<colloc>`, `<def>`, `<dictScrap>`, `<etym>`, `<form>`, `<gen>`, `<gram>`, `<gramGrp>`, `<hom>`, `<hyph>`, `<iType>`, `<lang>`, `<lbl>`, `<mood>`, `<number>`, `<oRef>`, `<oVar>`, `<orth>`, `<pRef>`, `<pVar>`, `<per>`, `<pos>`, `<pron>`, `<re>`, `<sense>`, `<stress>`, `<subc>`, `<superEntry>`, `<syll>`, `<tns>`, `<usg>`, `<xr>`

```
<entry>
  <form>
    <orth>poussin</orth>
    <pron>[pusè]</pron>
  </form>
  <gramGrp>
    <pos>n.</pos>
    <gen>m.</gen>
  </gramGrp>
  <sense n="1">Jeune poulet, nouvellement sorti de l'œuf, encore couvert de duvet. La
    poule et ses poussins.</sense>
  <sense n="2">Zool. Jeune oiseau (par rapport aux adultes, aux parents). </sense>
  <sense n="3"> (êtres humains) <def n="1">Fam. Terme d'affection (enfant). </def>
    <def n="2"> Sports. Catégorie d'âge (9 ans) qui précède celle des benjamins.</def>
    <def n="3">Elève de première année dans certaines écoles (Air,
      Aéronautique).</def>
  </sense>
</entry>
```

10. Manuscript Description

- Ce module fournit un élément `<msDesc>`, disponible ou dans un `teiHeader`, ou dans un texte, pour encoder une description structurée d'un manuscrit, d'un incunable, ou d'un autre objet inscrit.
- Il est organisé en 4 parties principales : `<msIdentifier>`, `<msContents>`, `<physDesc>`, `<history>`, `<additional>`, suivi facultativement de plusieurs `<msPart>`.
- Ses composants peuvent contenir des paragraphes en prose, ou encore des éléments plus spécifiques.

```
<msDesc>
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Add. A. 61</idno>
    <altIdentifier type="SC">
      <idno>28843</idno>
    </altIdentifier>
  </msIdentifier>
  <p>In Latin, on parchment: written in more than one hand of the 13 th cent. in England: 7¼ x 5¾ in., i + 55 leaves, in double columns: with a few coloured capitals.</p>
  <p>'Hic incipit Bruitus Anglie,' the De origine et gestis Regum Angliae of Geoffrey of Monmouth (Galfridus Monumetensis: beg. 'Cum mecum multa & de multis.'</p>
  <p>On fol. 54v very faint is 'Iste liber est fratris guillelmi de buria de ... Roberti ordinis fratrum Predicatorum' 14th cent. (?): 'banvillae' is written at the foot of
```

10. Manuscript Description (contd.)

Composants spécialisés : <accMat>, <acquisition>, <additions>, <adminInfo>, <altIdentifier>, <binding>, <bindingDesc>, <catchwords>, <collation>, <collection>, <colophon>, <condition>, <custEvent>, <custodialHist>, <decoDesc>, <decoNote>, <depth>, <dimensions>, <explicit>, <filiation>, <finalRubric>, <foliation>, <handDesc>, <height>, <heraldry>, <incipit>, <institution>, <layout>, <layoutDesc>, <locus>, <locusGrp>, <material>, <msItem>, <msName>, <musicNotation>, <objectDesc>, <origDate>, <origPlace>, <origin>, <provenance>, <recordHist>, <repository>, <rubric>, <seal>, <sealDesc>, <secFol>, <signatures>, <source>, <stamp>, <summary>, <support>, <supportDesc>, <surrogates>, <textLang>, <typeDesc>, <watermark>, <width>

11. Representation of Primary Sources

Fournit des éléments spécialisés pour le balisage des aspects physiques d'un document, complémentaires à ceux contenus dans le noyau.

- éléments de transcription: `<addSpan>`, `<am>`, `<damage>`, `<damageSpan>`, `<delSpan>`, `<ex>`, `<fw>`, `<handNotes>`, `<handShift>`, `<restore>`, `<space>`, `<subst>`, `<supplied>`
- éléments pour textes numérisés en mode-image: `<facsimile>`, `<surface>`, `<zone>`

```
<facsimile>  
  <graphic url="page1.png"/>  
  <graphic url="page2.png"/>  
  <graphic url="page3.png"/>  
  <graphic url="page4.png"/>  
</facsimile>
```

12. Critical Apparatus

Éléments spécialisés pour la représentation de l'apparat critique classique : `<app>`, `<lacunaEnd>`, `<lacunaStart>`, `<lem>`, `<listWit>`, `<rdg>`, `<rdgGrp>`, `<variantEncoding>`, `<wit>`, `<witDetail>`, `<witEnd>`, `<witStart>`, `<witness>`

```
<app>
  <lem wit="wit1">comte</lem>
  <rdgGrp type="orthographic">
    <rdg wit="#wit2">cante</rdg>
    <rdg wit="#wit3">contes</rdg>
  </rdgGrp>
</app>
```

13. Names, Dates, People, Places

- Éléments spécialisés pour la représentation des nominations et des entités nommées : <addName>, <affiliation>, <age>, <birth>, <bloc>, <climate>, <country>, <death>, <district>, <education>, <event>, <faith>, <floruit>, <forename>, <genName>, <geo>, <geogFeat>, <geogName>, <langKnowledge>, <langKnown>, <listEvent>, <listNym>, <listOrg>, <listPerson>, <listPlace>, <location>, <nameLink>, <nationality>, <nym>, <occupation>, <offset>, <org>, <orgName>, <persName>, <person>, <personGrp>, <place>, <placeName>, <population>, <region>, <relation>, <relationGrp>, <residence>, <roleName>, <settlement>, <sex>, <socecStatus>, <state>, <surname>, <terrain>, <trait>

13. Names, Dates, People, Places

```
<person xml:id="Ovi01" sex="1" role="poet">
  <persName xml:lang="la">Publius Ovidius Naso</persName>
  <birth when="-0044-03-20"> 20 March 43 BC<placeName>
    <settlement type="city">Sulmona</settlement>
    <country key="IT">Italie</country>
  </placeName>
</birth>
<death notBefore="0017" notAfter="0018">17 or 18 AD<placeName>
  <settlement type="city">Tomis (Constanta)</settlement>
  <country key="RO">Roumanie</country>
</placeName>
</death>
</person>
<!-- ... -->
<p>Les Métamorphoses d'<name ref="#Ovi01">Ovide</name>, un long poème en quinze livres...
</p>
```


14. Tables, Formulae, and Graphics

- Éléments spécialisés pour les tables, les formulaires, etc :
<cell>, <figDesc>, <figure>, <formula>, <row>, <table>

```
<formula notation="TeX">$e=mc^2$</formula>
```

```
<table rows="4" cols="4">  
<thead>Persistence de la neige dans les Alpes suisses </thead>  
<tbody>  
<tr>  
<td role="label">A l'altitude de</td>  
<td role="data">650 m.</td>  
<td role="data">1300m.</td>  
<td role="data">1950m.</td>  
<td role="data">2700m.</td>  
</tr>  
<tr>  
<td role="label">la neige reste</td>  
<td role="data">77 jours.</td>  
<td role="data"> 200 jours.</td>  
<td role="data"> 245 jours.</td>  
<td role="data"> 365 jours.</td>  
</tr>  
</tbody>  
</table>
```

15. Language Corpora

Ce chapitre traite de la construction des grands corpus linguistiques, et par conséquent se réfère à presque tous les autres chapitres. Il propose d'ailleurs quelques éléments spécialisés pour l'enrichissement des métadonnées associées à un corpus :

- `<activity>`, `<channel>`, `<constitution>`, `<derivation>`,
`<domain>`, `<factuality>`, `<interaction>`, `<locale>`,
`<particDesc>`, `<preparedness>`, `<purpose>`, `<setting>`,
`<settingDesc>`, `<textDesc>`

```
<settingDesc>  
  <setting who="#p1 #p2">  
    <name> Clermont-Ferrand</name>  
    <date>Mi-décembre</date>  
    <locale>Ceyrat</locale>  
    <activity>mise en marche de la voiture</activity>  
  </setting>  
</settingDesc>
```

16. Linking, Segmentation, and Alignment

Ce chapitre fournit des éléments génériques de

- structuration : `<ab>`, `<seg>`
- liaison (hyperliens) : `<alt>`, `<altGrp>`, `<anchor>`, `<join>`,
`<joinGrp>`, `<link>`, `<linkGrp>`
- synchronisation : `<timeline>`, `<when>`, et plusieurs attributs.
- et les attributs associés : `@corresp`, `@synch`, `@sameAs`, `@next`,
`@prev`, `@exclude`, `@select`.

```
<linkGrp type="translation">
  <link targets="#CCS1 #SW1"/>
  <link targets="#CCS2 #SW2"/>
<!-- ... -->
</linkGrp>
<div type="volume" xml:id="CCS" xml:lang="fr">
  <p>
    <s xml:id="CCS1">Longtemps, je me suis couché de bonne heure.</s>
    <s xml:id="CCS2">Parfois, à peine ma bougie éteinte, mes yeux se fermaient si vite que
      je n'avais pas le temps de me dire : "Je m'endors."</s>
  </p>
<!-- ... -->
</div>
<div type="volume" xml:id="SW" xml:lang="en">
  <p>
    <s xml:id="SW1">For a long time I used to go to bed early.</s>
    <s xml:id="SW2">Sometimes, when I had put out my candle, my eyes would close so
      quickly that I had not even time to say "I'm going to sleep."</s>
  </p>
</div>
```

17. Simple Analytic Mechanisms

Propose des éléments spécifiques pour les besoins communs de :

- segmentation linguistique: `<c>`, `<cl>`, `<m>`, `<phr>`, `<s>`, `<w>`
- analyses ou interprétations thématiques (etc.) `<interp>`, `<interpGrp>`, ``, `<spanGrp>`
- fournit en plus l'attribut `@ana`

```
<l>
  <w ana="#03">Le </w>
  <w ana="#SM">vin </w>
  <w ana="#V">sait </w>
  <w ana="#V">revêtir </w>
  <w ana="#A">le plus </w>
  <w ana="#J">sordide </w>
  <w ana="#SM">bouge</w>
</l>
<l>
  <w ana="#DF">D' </w>
  <w ana="#DF">un </w>
  <w ana="#SM">luxe </w>
  <w ana="#J">miraculeux</w>
  <c ana="#CC">,</c>
</l>
<l>
  <w ANA="#CC">Et </w>
  <w ana="#V">fait </w>
  <w ana="#V">surgir </w>
  <w ana="#0">plus </w>
  <w ana="#DF">d' </w>
  <w ana="#DF">à </w>
```

18. Feature Structures

Ce chapitre décrit un jeu de balises pour représenter un système analytique très répandu dans la linguistique théorique et maintenant standardisé au niveau ISO. La sémantique est modélisée en terme de *traits* nommés et leurs valeurs, qui peuvent être regroupés dans des "systèmes de traits" quasiment formelles. Cette norme permet l'interopérabilité des lexiques et des autres systèmes de gestion d'intelligence artificielle. Il y a deux parties :

- représentations des traits : `<binary>`, `<f>`, `<fs>`, `<fsDecl>`, `<fsDescr>`, `<numeric>`, `<string>`, `<symbol>`, `<vAlt>`, `<vColl>`, `<vDefault>`, `<vLabel>`, `<vMerge>`, `<vNot>`, `<vRange>`
- déclaration de système de traits : `<bicond>`, `<cond>`, `<default>`, `<fDecl>`, `<fDescr>`, `<fLib>`, `<fsConstraints>`, `<fsdDecl>`, `<fsdLink>`, `<fvLib>`, `<if>`, `<iff>`, `<then>`,

18. Feature Structures (contd.)

```
<fs type="real_estate_listing">
  <f name="selling.points">
    <vColl org="set">
      <string>alarm system</string>
      <string>good view</string>
      <vAlt>
        <string>pool</string>
        <string>jacuzzi</string>
      </vAlt>
    </vColl>
  </f>
</fs>
```

19. Graphs, Networks, and Trees

Ce chapitre propose des éléments pour la représentation de quelques structures abstraites, notamment les arborescences, les réseaux, et les graphes.

Sont définis : `<arc>`, `<eLeaf>`, `<eTree>`, `<forest>`, `<forestGrp>`, `<graph>`, `<iNode>`, `<leaf>`, `<node>`, `<root>`, `<tree>`, `<triangle>`

```
<graph>
  <node xml:id="LAX2" degree="2" adj="#LVG2 #PHX2">
    <label>LAX2</label>
  </node>
  <node xml:id="LVG2" degree="2" adj="#LAX2 #PHX2">
    <label>LVG2</label>
  </node>
<!-- ... -->
</graph>
```

20. Non-hierarchical Structures

Ce chapitre fait le bilan de plusieurs techniques proposées pour le balisage des structures non-imbriquées / hiérarchies superposées, notamment les méthodes "stand-off" (débordé).

Il ne propose pas de balises spécifiques.

21. Certainty and Responsibility

- Éléments et autres mécanismes pour indiquer le degré de certitude, la précision, ou la responsabilité envers l'encodage : `<certainty>`, `<respons>` and `<precision>`

```
Elisabeth aimait bien <persName xml:id="ESX">Essex</persName>
<!-- ... -->
<certainty target="#ESX" locus="gi" degree="0.6">
  <desc>réfère probablement à l'Earl de Essex</desc>
</certainty>
<certainty target="#CE-
pl1" locus="gi" degree="0.4" assertedValue="placeName">
  <desc>pourrait référer au comté de Essex</desc>
</certainty>
```

22. Documentation Éléments

Ce chapitre définit des éléments spécialisés pour l'encodage des manuels d'encodage, notamment ceux qui se servent de la TEI.

- Éléments appropriés à tout système d'encodage : `<code>`, `<ident>`, `<att>`, `<gi>`, `<tag>`, `<val>`, `<eg>`, `<egXML>`.
- Éléments spécialisés pour le système ODD: `<schemaSpec>`, `<elementSpec>`, `<moduleRef>`, etc.

```
<schemaSpec prefix="TEI_" ident="testsvg" start="TEI svg">
  <moduleRef key="header"/>
  <moduleRef key="core"/>
  <moduleRef key="tei"/>
  <moduleRef key="textstructure"/>
  <moduleRef url="svg11.rng"/>
</schemaSpec>
```

23. Using the TEI

Ce chapitre rassemble des informations sur l'utilisation du système TEI, en particulier :

- comment personnaliser le schéma TEI
- définition formelle de *TEI conformance*
- implémentation d'un processeur ODD