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



I18n

Suite à un effort financé par le ALLC (entre autres), les définitions, les glosses, et quelques exemples sont traduits en plusieurs langues, notamment le français:

... mais les chapitres des Guidelines actuels restent jusqu'à présent anglophone.



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



Et, en matières annexes ...

- une bibliographie
- définition de référence pour tout
 - élément, plus ses attributs locals
 - classe d'attribut, avec les attributs qu'elle fournit
 - macro (patron)
 - macro (datatype)

http://www.tei-c.org/Guidelines/REFTAG.html"



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

- 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. Elements 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, ette 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.



2. The TEI Header

• Éléments proposés : <applifo>, <application>, <authority>, <availability>, <biblFull>, <cRefPattern>, <calendar>, <calendarDesc>, <catDesc>, <catRef>, <category>, <change>, <classCode>, <classDecl>, <correction>, <creation>, <distributor>, <edition>, <editionStmt>, <editorialDecl>, <encodingDesc>, <extent>, < fileDesc>, <funder>, <geoDecl>, <handNote>, <hyphenation>, <idno>, <interpretation>, <keywords>, <langUsage>, <language>, licence>, <listChange>, <namespace>, <normalization>, <notesStmt>, <principal>, <profileDesc>, <projectDesc>, <publicationStmt>, <quotation><refState>, <refsDecl>, <rendition>, <revisionDesc>, <samplingDecl>, <scriptNote>, <segmentation>, <seriesStmt>, <sourceDesc>, <sponsor>, <stdVals>, <tagUsage>, <tagsDecl>, <taxonomy>, <teiHeader>, <textClass>, <titleStmt>, <typeNote>



<teiHeader> exemple

```
<teiHeader>
<fileDesc>
<titleStmt>
<title>Titre: un document TEI</title>
</titleStmt>
<publicationStmt>
Informations sur la publication
</publicationStmt>
<sourceDesc>
Informations sur ses sources
</sourceDesc>
</fileDesc>
</teiHeader>
```



3. Elements Available in All TEI Documents

• Éléments proposés :<abbr>, <add>, <addrLine>, <address>, <analytic>, <author>, <bibl>, <biblScope>, <biblStruct>, <binaryObject>, <cb>, <choice>, <cit>, <corr>, <date>, , <desc>, <distinct>, <divGen>, <editor>, <email>, <emph>, <expan>, <foreign>, <qap>, <qb>, <qloss>, <qraphic>, <head>, <headItem>, <headLabel>, <hi>, <imprint>, <index>, <item>, <l>, <label>, <lb>, <lg>, , , list>, , listBibl>, <measure>, <measureGrp>, <meeting>, <mentioned>, <milestone>, <monogr>, <name>, <note>, <num>, <orig>, , <pb>, <postBox>, <postCode>, <ptr>, <pubPlace>,</pr> <publisher>, <q>, <quote>, <ref>, <req>, <relatedItem>, <resp>, <respStmt>, <rs>, <said>, <series>, <sic>, <soCalled>, <sp>, <speaker>, <stage>, <street>, <teiCorpus>, <term>, <textLang>, <time>, <title>, <unclear>



4. Default Text Structure

Éléments proposés: <TEl>, <argument>, <back>, <body>,
 <byline>, <closer>, <dateline>, <div>, <div1>, <div2>,
 <div3>, <div4>, <div5>, <div6>, <div7>, <docAuthor>,
 <docDate>, <docEdition>, <docImprint>, <docTitle>,
 <epigraph>, <floatingText>, <front>, <group>, <imprimatur>,
 <opener>, <postscript>, <salute>, <signed>, <text>,
 <titlePage>, <titlePart>, <trailer>



5. Representation of Non-standard Characters and Glyphs

Éléments proposés : <char>, <charDecl>, <charName>,
 <charProp>, <g>, <glyph>, <glyphName>, <localName>,</mapping>, <unicodeName>, <value>



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>Le flot huileux et lourd décompose ses m<rhyme label="a">oires</rhyme> ; </l>
  <l>Les ul'océan blê<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>,</actor>, <castGroup>, <castItem>, <castList>, <epilogue>, <move>,, <performance>, <prologue>, <role>, <roleDesc>, <set>,, <sound>, <tech>, <view>

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



8. Transcriptions of Speech

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>)
- des éléments pour la structuration des enonces, notamment par rapport a un 'time line': ligne de temps



Exemple de text oral

```
<vocal iterated="true">
 <desc>sifflements intermittents</desc>
</vocal>
<u who="#spkr1" sync="#T1">si tu te déplaçais</u>
<u who="#spkr2" sync="#T2">Joe et moi l'aurions mis entre nous</u>
<!-- -->
listPerson>
 <person xml:id="spkr1">
  <persName>Jules</persName>
 </person>
 <person xml:id="spkr2">><persName>Jim</persName>
 </person>
</listPerson>
<!-- ... -->
<timeLine>
 <when xml:id="T1" absolute="12:34:45"/>
 <when xml:id="T2" absolute="12:35:05"/>
</timeLine>
```



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>, <sense>, <stress>, <subc>, <superEntry>, <syll>, <tns>, <usg>, <xr>

```
<entrv>
 <form>
   <orth>poussin</orth>
   on>[pusë]
 </form>
 <qramGrp>
   <pos>n.</pos>
   <qen>m.</qen>
 </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.



Exemple de manscript description

```
<msDesc>
 <msIdentifier>
   <settlement>Oxford</settlement>
   <repository>Bodleian Library/repository>
   <idno>MS. Add. A. 61</idno>
   <altIdentifier type="SC">
     <idno>28843</idno>
   </altIdentifier>
 </msIdentifier>
 In Latin, on parchment: written in more than one hand of the 13 th cent. in England:
    7\frac{1}{4} \times 5\frac{3}{8} in., i + 55 leaves, in double columns: with a few coloured capitals.
 'Hic incipit Bruitus Anglie,' the De origine et gestis Regum Angliae of Geoffrey of
    Monmouth (Galfridus Monumetensis: beg. 'Cum mecum multa & de multis.'
 On fol. 54v very faint is 'Iste liber est fratris guillelmi de buria de ... Roberti
    ordinis fratrum Pred[icatorum], ' 14th cent. (?): 'hanauilla' is written at the foot of
    the page (15th cent.). Bought from the rev. W. D. Macray on March 17, 1863, for £1
    10s.
</msDesc>
```



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>,
<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>, <am>,
 <ex> ...
- éléments pour textes numérisés en mode-image : <facsimile>,
 <surface>, <surfaceGrp>, <zone>
- éléments pour l'édition génétique : <sourceDoc>. <change>,
 <metamark>,
- et des attributs supplementaires notamment : @facs, @ulx, @uly, @lrx, @lry, @change, @resp, @cert ...



Exemple de transcription d'un manuscrit



```
<facsimile>
 <surface xml:id="p1">
   <qraphic url="page1.png"/>
   <graphic url="page1.tiff"/>
 </surface>
</facsimile>
<text>
 <ab facs="#p1">
   <lb/>
   <hi rend="lettrine">H</hi>
   <q ref="#wvnn">W</q>ÆT WE GARDE
 <lb/>na in gear-dagum beod-cyninga
 <lb/>lb/>prym gefrunon, hu ða æbelingas
 <lb/>ellen fremedon. oft scyld scefing sceape<add>na</add>
   <lb/>breatum. moneq<ex>um</ex> mæqbum meodo-setl<add>a</add>
   <lb/>of<damage>
     <desc>tache d'encre</desc>
   </damage>teah ...
 </ab>
</text>
```



12. Critical Apparatus

Éléments spécialisés pour la représentation de l'apparat critique classique : <app>, <lacunaEnd>, <lacunaStart>, <lem>, tistWit>, <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>
</dgGrp>
```



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>, <geoqFeat>, <geoqName>, <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>



Exemple de balisage des infos personnels

```
<person xml:id="0vi01" sex="1" role="poet">
        <persName xml:lang="la">Publius Ovidius Naso</persName>
        <br/>

                          <settlement type="city">Sulmona</settlement>
                          <country kev="IT">Italie</country>
                 </placeName>
        </birth>
        <death notBefore="0017" notAfter="0018">17 or 18 AD<placeName>
                          <settlement type="city">Tomis (Constanta)</settlement>
                          <country key="R0">Roumanie</country>
                 </placeName>
        </death>
</person>
<!-- ... -->
Les Métamorphoses d'<name ref="#0vi01">0vide</name>, un long poème en quinze livres...
```



14. Tables, Formulae, and Graphics

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

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

```
<head>Persistance de la neige dans les Alpes suisses </head>
 <row>
   <cell role="label">A l'altitude de</cell>
   <cell role="data">650 m.</cell>
   <cell role="data">1300m.</cell>
   <cell role="data">1950m.</cell>
   <cell role="data">2700m.</cell>
 </row>
 <row>
   <cell role="label">la neige reste</cell>
   <cell role="data">77 jours.</cell>
   <cell role="data"> 200 iours.</cell>
   <cell role="data"> 245 iours.</cell>
   <cell role="data"> 365 jours.</cell>
 </row>
```



15. Language Corpora

Ce chapitre traite de la construction des grands corpus linguistiques, et par conséquence 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>,</dorantory, <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">
 targets="#CCS1 #SW1"/>
 targets="#CCS2 #SW2"/>
<!-- ... -->
<div type="volume" xml:id="CCS" xml:lang="fr">
 >
   <s xml:id="CCS1">Longtemps. ie 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>
 <!-- -->
</div>
<div type="volume" xml:id="SW" xml:lang="en">
   <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>
```

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

```
<1>
 <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>
</1>
<1>
 <w ana="#DF">D'</w>
 <w ana="#DF">un </w>
 <w ana="#SM">luxe </w>
 <w ana="#J">miraculeux</w>
 <c ana="#CC">,</c>
</1>
<1>
 <w ANA="#CC">Ft </w>
 <w ana="#V">fait </w>
 <w ana="#V">surgir </w>
 <w ana="#0">plus </w>
     ana="#DF">d'</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">
</cold org="set">
<string>alarm system</string>
<string-good view</string>
<vAlt>
<string>pool</string>
<string>jacuzzi</string>
</vAlt>
</vColl>
</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>



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>



22. Documentation Elements

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



Quelques liens utils

- Guidelines: http://www.tei-c.org/Guidelines/P5/
- Référence: http://www.tei-c.org/release/doc/tei-p5doc/en/html/REF-ELEMENTS.html
- Badges: http://www.tei-c.org/About/Badges/
- Wiki: http://wiki.tei-c.org/
- Sourceforge: http://sourceforge.net/projects/tei/

