

Atelier XML TEI

Ariane Pinche ¹², Matthias Gille Levenson³²

¹CNRS, ²CIHAM, ³ÉNS Lyon

24 janvier 2023

Table of Contents

- 1 Éléments de base en TEI
- 2 Les TEI guidelines
- 3 Structurer un index
- 4 Quelques exemples d'encodage courant

Structuration générale d'un fichier XML TEI

Tout document TEI a au moins deux parties :

- Un en-tête, représenté au moyen d'un élément `<teiHeader>` contenant des métadonnées décrivant le document ;
- le texte lui-même, représenté par un élément `<text>` qui peut être subdivisé en sous-unités.

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- métadonnées décrivant un texte -->
  </teiHeader>
  <text>
    <!-- texte -->
  </text>
</TEI>
```

Structuration du `teiHeader`

Le **teiHeader** minimal comporte au minimum un élément *fileDesc* contenant les trois sections suivantes :

- un élément **titleStmt** : informations identifiant le document lui-même
- un élément **publicationStmt** : informations sur la façon dont il est distribué ou publié
- un élément **sourceDesc** : indications sur ses origines

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Titre de l'œuvre</title>
    </titleStmt>
    <publicationStmt>
      <p>Informations sur la publication de l'œuvre.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informations sur la source du texte</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Structuration du corps du texte

l'élément **text** peut contenir les éléments suivants :

- un élément **front** pour le texte liminaire (avant-propos, préface, etc..)
- un élément **body** pour le corps de texte
- un élément **back** pour les annexes, postface, etc.

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- métadonnées décrivant un texte -->
  </teiHeader>
  <text>
    <front/>
    <body/>
    <back/>
  </text>
</TEI>
```

Principaux éléments de structuration du texte

L'information textuelle peut être hiérarchisée au moyen des éléments suivants

- **div** pour les grandes divisions ou subdivisions du texte
- **p** pour les paragraphes
- **lg** pour les strophes
- **l** pour les vers
- **sp** pour les répliques
- Ces éléments peuvent être personnalisés avec les attributs *type* et *n*

```
<text>
  <body>
    <div type="book" n="1">
      <div type="chapter" n="1">
        <p>Some text here.</p>
      <p/>
    </div>
  </div>
</body>
</text>
```

Structurer en XML TEI le poème de Paul Verlaine, *Mon rêve Familier*

Lien vers le texte : https://fr.wikipedia.org/wiki/Mon_rêve_familier

- Ouvrir Oxygen Editor
- Ouvrir un nouveau fichier *XML TEI all*
- Remplir un `teiHeader` minimal
- Copier le texte de Verlaine dans le **body**
- Structurer le poème en vous aidant des éléments vus précédemment

Table of Contents

- 1 Éléments de base en TEI
- 2 Les TEI guidelines**
- 3 Structurer un index
- 4 Quelques exemples d'encodage courant

Les recommandations visent à :

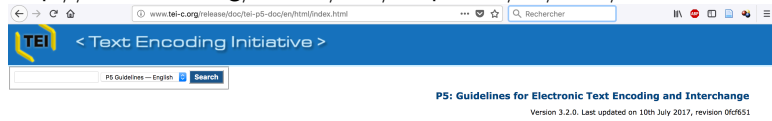
- Fournir un format standard
- Favoriser l'échange de textes
- Proposer des ensembles de conventions d'encodage adaptés à des applications différentes

TEI guidelines, mode d'emploi

Comment lire les guidelines ?

La page d'accueil :

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>



Front Matter

- [Title](#)
- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

Text Body

- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Characters, Glyphs, and Writing Modes](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Dictionaries](#)
- 10 [Manuscript Description](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulae, Graphics and Notated Music](#)
- 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, Precision, and Responsibility](#)
- 22 [Documentation Elements](#)
- 23 [Using the TEI](#)

TEI sourcecode

- [Getting and Using the TEI Sources](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

Comment lire les spécifications d'élément ?

www.tei-c.org/release/doc/tei-p5-doc/html/ref-msIdentifier.html

Rechercher

Version 3.2.0. Last updated on 10th July 2017, revision 06cf651

<msIdentifier>

Accueil
C Éléments

<msIdentifier> (identifiant du manuscrit) Contient les informations requises pour identifier le manuscrit en cours de description. [10.4 The Manuscript Identifier]

Module	msdescription — Manuscript Description	L'élément est documenté dans le module msdescription (10 Manuscript Description)
Attributs	<div><div>att.global</div><div>@xml:id, @n, @xml:lang, @xml:base, @xml:space (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @include)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) (att.global.source (@source))</div><div>attribute class</div></div>	
Membre du	model.bibliPart	model class
Contenu dans	<div>core: bibli</div> <div>msdescription: msDesc msFrag msPart</div> <div>éléments regroupés par modules</div>	
Peut contenir	<div>header: idno</div> <div>msdescription: altIdentifier collection institution msName repository</div> <div>namesdates: bloc country district geoName placeName region settlement</div>	
Exemple	<pre><msIdentifier> <country>France</country> <settlement>Paris</settlement> <repository xml:lang="fr">Bibliothèque nationale de France. Réserve des livres rares</repository> <idno>B-73</idno> <!-- dans le cas des recueils : cote uniquement sans les sous-cotes --> <altIdentifier> <idno>B-121</idno> <note>Cote de la bibliothèque royale au XVIIIe siècle (inscrite à l'encre, sur la doubleure de table).</note> </altIdentifier> <altIdentifier> <idno>Double de B. 274. A (Réserve)</idno> <note>Cote inscrite face à la page de titre, en remplacement de la cote "1541", barrée</note> </altIdentifier> </msIdentifier></pre> <div>Toute la liste bibliography</div>	

En vous aidant des TEI guidelines :

- Trouver la documentation pour encoder des répliques de pièce de théâtre
- Comment peut-on indiquer qui prononce la réplique ?

Table of Contents

- 1 Éléments de base en TEI
- 2 Les TEI guidelines
- 3 Structurer un index**
- 4 Quelques exemples d'encodage courant

Structuration d'un index

En XML TEI, on peut signaler des noms de lieu ou de personne et même commencer à structurer un index des personnages ou des lieux au fil de l'édition d'un texte.

Les entités nommées sont signalées à l'aide des balises *persName* ou *placeName*.

```
<text n="edition">
  <body n="urn:cts:froLit:jns915.jns2117.ciham-fro1">
    <div n="1"
      type="chapter"
      facs="http://gallica.bnf.fr/ark:/12148/btv1b84259980/f237, http://gallica.bnf.fr/ark:/12148/btv1b84259980/f238"
      corresp="urn:cts:latinLit:stoa0270.stoa003:1.1, urn:cts:latinLit:stoa0270.stoa003:1.2">
      <head>Ci commence<lb/>li dialogues que <persName ref="#postumien">postumiens</persName> <persName ref="#gallus">Gal
      <persName ref="#severus">seuerus</persName> le
      rate<lb/> les oi <app>
        <lem>ler</lem>
        <rdg wit="#C2 #C3" type="ajout">parler et qi i fu</rdg>
      </app>
    </head>
    <div n="1" type="section">
      <p>
        <pb n="114" facs="http://gallica.bnf.fr/ark:/12148/btv1b84259980/f237"/>
        <hi rend="rubricated orig">Ci commence<lb/>li dialogues que <persName ref="#postumien">postumiens</persName>
        de <persName ref="#martin">seint martin</persName> des
        moines d<placeName ref="#egypte" full="yes">egy<lb/>pte</placeName> si <persName ref="#severus">seuerus</persName>
        oi ler</hi>
        <lb/>
        <hi rend="decorated-initial">U</hi>N ior auint<lb/>qeie
      <persName ref="#gallus">gallus</persName>
      <lb/>
      <cb n="b"/>
    </div>
  </body>
</text>
```

Structuration d'un index

Pour que toutes les occurrences d'une entité puissent pointer vers une entrée unique, il faut la déclarer soit dans l'élément *profilDesc* du *teiHeader* soit dans un élément *standOff* dans une liste de personnes ou de lieux.

```
<listPerson>
  <person xml:id="agaridut">
    <persName>Agariduz</persName>
    <note type="biographical">Honnête homme, témoin de l'adoucissement des mœurs cruelles
d'Avicien.</note>
  </person>
  <person xml:id="agnes">
    <persName>Agnies</persName>
    <death when="0300">4<ex>e</ex> siècle</death>
    <note type="biographical" source="http://data.bnf.fr/11958468/agnes/">Martyre
romaine.</note>
  </person>
  <person xml:id="antoine">
    <persName>Antoines</persName>
    <birth when="0251">251</birth>
    <death when="0356">356</death>
    <note type="biographical" source="http://data.bnf.fr/11888967/antoine/">
      <surname type="complex" full="yes">Antoine le grand, Antoine d'Égypte ou Antoine
l'Ermite</surname>, considéré comme le fondateur du l'érémisme chrétien.</note>
  </person>
  <person xml:id="arborius">
    <persName>Arborius</persName>
    <note type="biographical">Magnus Arborius, élu comte des largesses sacrées en 379, puis
préfet de Rome en 380, il aurait été l'un des protecteurs de <persName ref="#saintMartin">saint Martin</persName>.</note>
  </person>
  <person xml:id="arpage">
    <persName>Arpage</persName>
    <note type="biographical">Prêtre.</note>
  </person>
```

Documentation:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

Mise en pratique : Structurer un index

Le lien entre l'index et les occurrences grâce à un jeu d'identifiants et de pointeurs.

- L'identifiant est unique, il est déclaré dans la liste des entités de l'index avec l'attribut *xml:id*
- ex : `<person xml:id="agaridut">`
- Le pointeur est ajouté sur les occurrences de l'entité dans le texte avec l'attribut *ref*
- ex : `<persName ref="#agaridut">`
- la valeur du pointeur est la chaîne de caractère de id précédée d'une dièse.

Encoder les deux premiers paragraphes de *Notre Dame de Paris*, livre II, chapitre 1, «De Charybde en Scylla » (début de chapitre à « La cohue admirait »).

- Structurer le texte en livre, chapitre et paragraphes
- Baliser les noms de lieux
- Réaliser un index

Lien vers le texte : https://fr.wikisource.org/wiki/Notre-Dame_de_Paris/Livre_deuxième#II._La_place_de_Gr.C3.A8ve

Mise en pratique : Structurer une pièce de théâtre

Convertir en XML TEI, un extrait du Misanthrope de Molière en vous aidant de la documentation et des balises suivantes : `<castList>`, `<castItem>`, `<speaker>`, `<stage>`, `<head>`, `<div>`, `<sp>` et `<l>`.

Lien vers le texte :

https://fr.wikisource.org/wiki/Le_Misanthrope/Édition_Louandre,_1910

Documentation dans les guidelines :

<https://tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>

Table of Contents

- 1 Éléments de base en TEI
- 2 Les TEI guidelines
- 3 Structurer un index
- 4 Quelques exemples d'encodage courant

Structuration d'un encodage "paleographique"

Parmi les nombreuses possibilités qu'offre la TEI, on peut produire un **encodage imitatif** du manuscrit, et par exemple, signaler les abréviations tout en proposant leur développement, consigner une graphie fautive et proposer une correction. Ces propositions alternatives peuvent être rendues au fil du texte grâce à la balise *choice*.

Quelques balises courantes

- Développements d'abréviation : `<choice><abbr></abbr>
<expan></expan></choice>` ;
- Régularisations : `<choice><orig></orig><reg></reg>
</choice>` ;
- Corrections : `<choice><sic></sic><corr></corr></choice>` .

Documentation sur la représentation des sources primaires :
<http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/PH.html>

Structuration d'un appareil critique

XML TEI permet également de constituer des appareils critiques. Il existe plusieurs méthodes pour encoder les appareils, dont la plus simple est la *parallel segmentation method* qui permet d'ajouter l'apparat directement sur le texte du manuscrit de base.

```
<app>
  <lem wit="#El #Hg">Experience</lem>
  <rdg wit="#La" type="substantive">Experiment</rdg>
  <rdg wit="#Ra2" type="substantive">Eryment</rdg>
</app>
```

Documentation sur la représentation des appareils :

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPLK>

Structuration d'un appareil critique

Pour pouvoir lier les leçons aux témoins, il faudra au préalable déclarer dans le *teiHeader* la liste des témoins.

Déclaration des sources dans le `<sourceDesc>` :

- `<listWit>` : donne une liste de tous les témoins cités dans un appareil critique;
- `<witness>` : témoin d'une édition.

```
<listWit>
  <witness xml:id="C1">Paris, BnF, ms 412</witness>
  <witness xml:id="C2">Paris, BnF, ms 411</witness>
  <witness xml:id="C3">London, British Library,
    royal 20 D. VI.</witness>
</listWit>
```

Structuration de données linguistiques

Les éditions numériques peuvent être enrichies avec de nombreuses informations qu'il est impossible d'ajouter sur une édition papier. Parmi ces possibilités, l'ajout d'annotations linguistiques qui permettront par la suite l'étude de la langue du texte.

```
<w xml:id="t78"
  n="78"
  lemma="cel"
  pos="PR0dem"
  msd="NOMB.=p|GENRE=m|CAS=r">ceus</w>
<w xml:id="t79"
  n="79"
  lemma="venir"
  pos="VERcjc"
  msd="MODE=ind|TEMPS=psp|PERS.=3|NOMB.=s">vint</w>
<w xml:id="t80" n="80" lemma="por2" pos="PRE">por</w>
<w xml:id="t81"
  n="81"
  lemma="nostre"
  pos="DETpos"
  msd="PERS.=1|NOMB.=s|GENRE=m|CAS=r">nostre</w>
<w xml:id="t82"
  n="82"
  lemma="seigneur"
  pos="NOMcom"
  msd="NOMB.=s|GENRE=m|CAS=r">Signor</w>
<w xml:id="t83"
  n="83"
  lemma="öir"
  pos="VERinf"
  msd="NOMB.=x|GENRE=x|CAS=x">oïr</w>
```

Documentation sur l'annotation linguistique :

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/A1.html#AILALW>