### Séance 5

# Les TEI Guidelines

Modules, modèles, macros, etc.

# Correction

**Encodage d'un <u>TEI</u> header** 

# 2023-2024 | Devoirs XML-TEI M2 TNAH

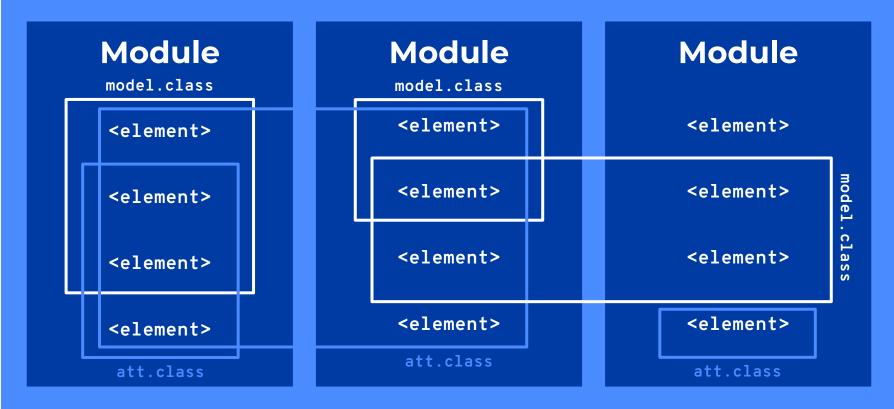
Les Guidelines guident les pratiques d'encodage, mais ne sont pas un standard qui contraint ce que l'on doit faire

Les guidelines permettent de bénéficier de l'expérience accumulée en matière de stratégie d'encodage, de méthodologie

# Structure des guidelines

Comment sont construites les recommandations

## TEI



### **Comment lire les Guidelines?**

### **Exemples**

Pour mieux comprendre ce qu'il faut mettre dans chaque balise

### **Modules**

Lire la documentation en fonction du contenu à encoder

### Élément

Pour aller voir le modèle, le contenu et contenant d'une balise

### Classes

Pour comprendre plus en détails comment est structurée la TEI

#### **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

#### **Back Matter**

- Appendix B Attribute Classes
- ⊕ Appendix C Elements
- Appendix D Attributes
- ★ Appendix E Datatypes and Other Macros
- Appendix F Bibliography
- Appendix G Deprecations
- Appendix H <u>Prefatory Notes</u>

  Appendix I Colombon

#### Appendix I Colophon

### **Text Body**

- **3** Elements Available in All TEI Documents

- 7 Performance Texts
- 8 Transcriptions of Speech
- ⊕ 9 Dictionaries
- **±** 10 Manuscript Description
- 11 Representation of Primary Sources
- 13 Names, Dates, People, and Places
- 14 Tables, Formulæ, Graphics, and Notated Music
- **15 Language Corpora**
- 16 Linking, Segmentation, and Alignment
- **18 Feature Structures**
- 19 Graphs, Networks, and Trees
- 20 <u>Non-hierarchical Structures</u>
- **22 Documentation Elements**

#### **TEI sourcecode**

- Getting and Using the TEI Sources.
- TEI GitHub Repository
- · Bug Reports, Feature Requests, etc.

### **Préambules**

#### **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. <u>Languages and Character Sets</u>

#### **Back Matter**

- ★ Appendix A Model Classes
- Appendix B Attribute Classes
- ⊕ Appendix C Elements
- Appendix D Attributes
- Appendix F Bibliography
- Appendix H <u>Prefatory Notes</u>

Appendix I Colophon

Liste de l'intégralité des éléments, classes, attributs, type de données

### **Text Body**

- 3 Elements Available in All TEI Documents
- 4 Default Text Structure
- 5 Characters, Glyphs, and Writing Modes
- 7 Performance Texts
- 8 Transcriptions of Speech
- 9 Dictionaries
- 10 Manuscript Description
- 11 Representation of Primary Sources
- 13 Names, Dates, People, and Places
- 14 Tables, Formulæ, Graphics, and Notated Music
- **15 Language Corpora**
- 16 <u>Linking, Segmentation, and Alignment</u>
- 18 Feature Structures
- 19 Graphs, Networks, and Trees
- 20 Non-hierarchical Structures
- 21 <u>Certainty, Precision, and Responsibility</u>
- **22 Documentation Elements**
- 23 <u>Using the TEI</u>

#### **TEI sourcecode**

- Getting and Using the TEI Sources.
- TEI GitHub Repository
- · Bug Reports, Feature Requests, etc.

# Documentation des modules

### Les éléments

Chacun des éléments est documenté dans une page dédiée renseignant sur :

Définition	Module
Contenu	Contenant
Classes	Exemples

Home C Elements

<div> (text division) contains a subdivision of the front, body, or back of a text. [4.1 Divisions of the Body]</div>	
Module	textstructure — Default Text Structure
Attributes	att.global (@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, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change) (att.global.responsibility (@cert, @resp)) (att.global.source (@source)) att.divLike (@org, @sample) (att.metrical (@met, @real, @rhyme)) (att.fragmentable (@part)) att.typed (@type, @subtype) att.declaring (@decls) att.written (@hand)
Member of	model.divLike
Contained by	textcrit: lem rdg textstructure: back body div front
May contain	analysis: interp interpGrp span spanGrp certainty: certainty, precision respons core: bibl biblStruct cb cit desc divGen ellipsis gap gb head index I label lb lg list listBibl meeting milestone note noteGrp p pb q quote said sp stage dictionaries: entry entryFree superEntry drama: camera caption castList move sound spGrp tech view figures: figure notatedMusic table header: biblFull iso-fs: fLib fs fvLib linking: ab alt altGrp anchor join joinGrp link linkGrp timeline msdescription: msDesc namesdates: listEvent listNym listObject listOrg listPerson listPlace listRelation nets: eTree forest graph listForest tree
	spoken: annotationBlock incident kinesic pause shift u vocal writing

Home C Elements

<aiv> (text division</aiv>	o) contains a subdivision of the front, body, or back of a text. [4.1 Divisions of the Body]	
Module	textstructure — Default Text Structure module dans lequel est documenté l'élément	
Attributes		
Member of	model.divLike model class	
Contained by	textcrit: lem rdg textstructure: back body div front	
May contain	analysis: interp interpGrp span spanGrp certainty: certainty precision respons core: bibl biblStruct cb cit desc divGen ellipsis gap gb head index ! label lb !g list lis dictionaries: entry entryFree superEntry drama: camera caption castList move sound spGrp tech view figures: figure notatedMusic table header: biblFull iso-fs: fLib fs fvLib linking: ab alt altGrp anchor join joinGrp link linkGrp timeline msdescription: msDesc namesdates: listEvent listNym listObject listOrg listPerson listPlace listRelation nets: eTree forest graph listForest tree spoken: annotationBlock incident kinesic pause shift u vocal writing	éléments regroupés par module

### Trouver des exemples

#### **Example**

```
<body>
<div type="part">
 <head>Fallacies of Authority</head>
 The subject of which is Authority in various shapes, and the object, to repress all
    exercise of the reasoning faculty.
 <div n="1" type="chapter">
  <head>The Nature of Authority</head>
  With reference to any proposed measures having for their object the greatest
      happiness of the greatest number [...]
  <div n="1.1" type="section">
   <head>Analysis of Authority</head>
   What on any given occasion is the legitimate weight or influence to be attached to
        authority [...] 
  </div>
  <div n="1.2" type="section">
   <head>Appeal to Authority, in What Cases Fallacious.</head>
   Reference to authority is open to the charge of fallacy when [...] 
  </div>
 </div>
</div>
</body>
                                                                                            bibliography Show all &
```

# Exercice

Résoudre une problématique précise

## Consigne

En naviguant dans les *Guidelines* par les différents moyens évoqués, trouver quelle est la recommandation TEI pour l'encodage d'un retour à la ligne avec une **césure** et un **tiret** 

Et ceux qui étaient venus pour vendre du charbon et du blé vendent du charbon et du blé et de grandes îles entourées d'eau de tous côtés, de grandes îles avec des arbres à pneus et des piano métalliques bien stylés pour qu'on n'entende pas trop les cris des indigènes autour des plantations quand les colons facétieux essaient après dîner leur carabine à répétition.

Tentative de description d'un dîner de têtes à Paris-France, Jacques Prévert

### Les modules

Les modules sont un ensemble cohérent de balises regroupées pour répondre à une problématique précise.

### Type de document Type de données

Modules pour la poésie, le théâtre, les manuscrits, le discours oral, etc. Modules pour la définition d'un apparat critique, formules mathématiques, images, etc.

# Chaque module

- est **documenté** par un chapitre des *Guidelines*
- définit un certain nombre de composantes (éléments, classes, macros)
- pointe vers des **spécifications** techniques
- correspond à une **problématique** particulière
- peut être inclus ou non dans une personnalisation (à part les 4 obligatoires)

# **Modules obligatoires**

### <u>tei</u>

définition des classes, macros et types de données

### core

éléments disponibles dans tous les documents

### <u>header</u>

en-tête contenant les métadonnées du document

### **textstructure**

éléments de base pour structurer un texte de type livre

# **ODD TEI minimal**

```
<schemaSpec ident="TEI-minimal" start="TEI">
  <moduleRef key="tei"/>
  <moduleRef key="header"/>
  <moduleRef key="core"/>
  <moduleRef key="textstructure"/>
  </schemaSpec>
```

## Modules complémentaires

**analysis** Analysis and Interpretation <u>17 Simple Analytic Mechanisms</u>

certainty Certainty and Uncertainty 21 Certainty, Precision, and Responsibility

corpus Metadata for Language Corpora <u>15 Language Corpora</u>

**drama** Performance Texts <u>7 Performance Texts</u>

figures Tables, Formulae, Figures 14 Tables, Formulæ, Graphics, and Notated Music

gaijiCharacter and Glyph5 Characters, Glyphs, and Writing Modes

linkingLinking, Segmentation, and Alignment16 Linking, Segmentation, and Alignment

msdescription Manuscript Description <u>10 Manuscript Description</u>

namesdates Names, Dates, People, and Places <u>13 Names, Dates, People, and Places</u>

tagdocs Documentation Elements <u>22 Documentation Elements</u>

textcrit Text Criticism 12 Critical Apparatus

transcr Transcription of Primary Sources <u>II Representation of Primary Sources</u>

verse Verse <u>6 Verse</u>

# Exercice

Résoudre une problématique précise

## Consigne

En naviguant dans les *Guidelines* par les différents moyens évoqués, trouver trois différentes balises pouvant être utilisées pour encoder une prise de parole

### Les classes

Les classes sont définies dans un module mais peuvent porter sur des éléments d'un autre. Les éléments héritent des propriétés de la classe à laquelle ils appartiennent.

### att.class

Éléments qui partagent les mêmes attributs

### model.class

Éléments qui sont situés aux mêmes endroits dans le document

### att.class

Ensemble d'éléments qui <u>admettent les mêmes attributs</u> :

att.global Attributs pouvant être utilisé avec @xml:id, @n, @rend,

tout élément @xml:lang, etc.

att.linking Attributs pour lier les éléments entre @corresp, @next,

deux @prev,@sameAs,etc.

**att.ascribed** pour discours action pouvant être @who attribué à un individu spécifique

# Attributs d'une classe

#### att.written

Home

B Attribute Classes

Module	tei — The TEI Infrastruc	ture	
Members	att.damaged [damage damageSpan] att.textCritical [lem rdg rdgGrp] att.transcriptional [add addSpan del delSpan mod redo restore retrace rt subst substJoin undo] ab closer div figure fw head hi label line note noteGrp opener p path postscript salute seg signed stage text trailer zone		
Attributes	@hand 🕹	points to a <u>handNote</u> element describing the hand considered responsible for the content of the element concerned.	
		Status	Optional
		Datatype	teidata.pointer

### Attributs à retenir

### @xml:id

Pour attribuer un identifiant unique à l' élément qui porte l'attribut

### @type

Pour attribuer un type à l' élément qui porte l'attribut

### @n

Pour attribuer un nombre (par forcément unique) à l'élément

### @rend

Pour préciser un certain rendu graphique de l' élément dans la source ou la sortie

### model.class

Se définit par un <u>ensemble d'endroits où les membres de la</u> <u>classe sont autorisés</u> (...**Part**) mais implique également une forme de regroupement sémantique (...**Like**)

model.divPart Éléments qui peuvent être contenu par une div

model.phrase Éléments pouvant apparaître au niveau de la phrase

model.nameLike Regroupement sémantique pour les divers types de noms

# Modèle de classe

### model.rdgPart

Home A Model Classes

model.rdgPart group	s elements which mark the beginning or ending of a fragmentary manuscript or other witness. [12.1.5 Fragmentary Witnesses]
Module	textcrit — Critical Apparatus
Used by	<u>lem rdg</u>
Members	lacunaEnd lacunaStart wit witEnd witStart
Note	These elements may appear anywhere within the elements <u>lem</u> and <u>rdg</u> , and also within any of their constituent elements.

# Les macros et types de données

Ils définissent quel type de contenu peuvent être utilisés dans les éléments et les valeurs d'attributs

# Les <u>macros</u>

Les macros définissent des raccourcis pour désigner les valeurs attendues au sein des éléments/attributs, pour faire caractériser des :

- types de contenus standardisés (<u>macro.paraContent</u>);
- valeurs autorisées (<u>data.count</u>).

Elles sont utilisées dans les **modèles de contenu** pour faire référence à des séquences fréquemment rencontrés dans des éléments

# **Macros**

### macro.paraContent

Home

E Datatypes and Other Macros

Module	tei — The TEI Infrastructure
Used by	add camera caption case colloc corr damage def del docEdition emph gen gram hi hyph iType imprimatur lang lbl mod mood number orig orth per pos pron ref reg restore retrace rhyme salute secl seg sic signed sound stress subc supplied surplus syll tech title titlePart tns u unclear usowriting
Content model	
	<pre><content>   <alternate max0ccurs="unbounded" min0ccurs="0">     <textnode></textnode>     <classref key="model.paraPart"></classref>     </alternate>   </content></pre>
Declaration	XML synta
	macro.paraContent = ( text   model.paraPart )*

# Modèle de contenu

### macro.limitedContent

### <desc>

```
<content>
    <macroRef key="macro.limitedContent"/>
</content>
```

# Type de données

teidata.certainty

Home

	ty defines the range of attribute values expressing a degree of certainty.
Module	tei — The TEI Infrastructure
Used by	teidata.probCert Class:  • att.dimensions/@precision  Element:  • certainty/@cert • precision/@precision • purpose/@degree
Content model	<pre><content>   <vallist type="closed"></vallist></content></pre>
Declaration	teidata.certainty = "high"   "medium"   "low"   "unknown"
Note	Certainty may be expressed by one of the predefined symbolic values 'high', 'medium', or 'low'. The value 'unknown' should be used in cases where the encoder does not wish to assert an opinion

# Valeurs d'attributs

### teidata.certainty

### att.global.responsability

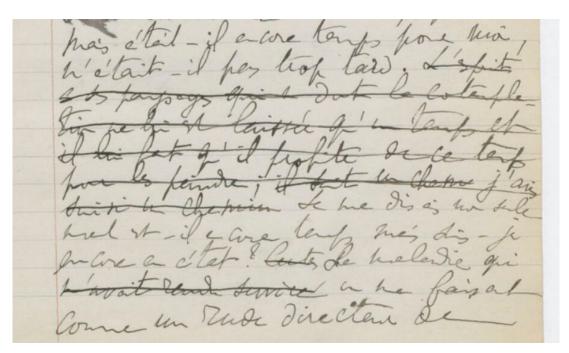
```
@cert (certainty) signifies the degree of certainty
Status Optional
Datatype teidata.probCert
```

# Exercice

Transcrire un passage raturé

## Consigne

En vous aidant du <u>chapitre sur les sources biffées</u>, transcrire ce passage des brouillons de <u>La recherche</u> à partir de <u>cette transcription</u>



BNF, Manuscrits, N. a. <u>Fr. 16727</u>, f. 117v

## Exercice

Baliser une liste de personnes

### Consigne

En vous aidant du <u>chapitre sur les noms de personnes et de lieux</u>, trouver comment baliser au sein du <particDesc>, du <settingDesc> et du <text> les mentions de personnes et lieux

La plupart des passagers embarqués à Brindisi avaient presque tous l'Inde pour destination. Les uns se rendaient à Bombay, les autres à Calcutta, mais via Bombay, car depuis qu'un chemin de fer traverse dans toute sa largeur la péninsule indienne, il n'est plus nécessaire de doubler la pointe de Ceylan.

Parmi ces passagers du *Mongolia*, on comptait divers fonctionnaires civils et des officiers de tout grade. De ceux-ci, les uns appartenaient à l'armée britannique proprement dite, les autres commandaient les troupes indigènes de cipayes, tous chèrement appointés, même à présent que le gouvernement s'est substitué aux droits et aux charges de l'ancienne Compagnie des Indes : sous-lieutenants à 7,000 francs, brigadiers à 60,000, généraux à 100,000.

Jules Verne, Le Tour du monde en quatre-vingts jours, Chapitre IX

# Ressources bibliographiques

Comment structurer une bibliographie

### Citer un livre?

Beaucoup de manières de faire!

### Différents niveaux de détails

**bibl** Référence bibliographique peu structurée dont les

sous-composants peuvent être ou non balisés

**biblFull** Référence bibliographique totalement structurée

biblStruct Une référence bibliographique structurée en sous-

éléments bibliographiques dans un ordre déterminé.

msDesc Description d'un unique manuscrit

**monogr** Sous-éléments bibliographiques à insérer dans

analytic biblStruct

## Référence bibliographique non-structurée

```
<bibl>Notre-Dame de Paris, Texte établi par Paul
Meurice, Librairie Ollendorff, 1904, [volume 1] [Section
A.] Roman, tome II. (p. 43-84)</bibl>
```

Méthode la plus simple (à éviter en général)

#### bibl

```
<bil>
    <title type="main">Notre-Dame de Paris</title>
    <title type="subtitle">Livre Deuxième</title>
    <author>
        <forename>Victor</forename>
        <surname>Hugo</surname>
    </author>
    <editor>Paul Meurice</editor>
    <publisher>Librairie Ollendorff</publisher>
    <distributor facs="https://fr.wiki...">
        Wikisource
    </distributor>
</bibl>
```

Pour une liste de références bibliographiques

## Exercice

Créer une bibliographie

## Consigne

Choisissez une ressource dans la <u>bibliographie de ce cours</u> et encodez les dans un élément stand0ff en vous appuyant sur les <u>recommandations des Guidelines</u>

# Désigner la source de l'encodage

publicationStmt et sourceDesc

## publicationStmt ou sourceDesc?

Méthode conditionnée par où l'on place notre document dans la chaîne éditoriale : est-ce qu'on considère notre encodage comme une édition en soi (sourceDesc) ou simplement comme une version d'un texte déjà publié (publicationStmt)?

# **Exemple Wikisource**



PARIS

IMPRIMÉ ÉDITÉ

PAR PAR

L'IMPRIMERIE NATIONALE LA LIBRAIRIE OLLENDORFF

MDCCCCIV

ŒUVRES COMPLÈTES DE VICTOR HUGO

ROMAN — II

NOTRE-DAME

**DE PARIS** 

## Exercice

Renseigner un sourceDesc

## Consigne

En vous appuyant sur les <u>recommandations</u> et les <u>exemples</u> fournis dans les <u>guidelines</u>, ainsi que sur les informations complémentaires sur l'édition disponibles dans <u>Gallica</u>, renseigner au maximum le <u>sourceDesc</u> pour une édition de <u>Notre Dame de Paris</u>

<bib>< . . . < / bibl>

### Création d'index

Les toponymes et anthroponymes

## Index générique

## Où placer un <u>index</u>?

#### profileDesc

Déclarations qui ne relèvent pas des données bibliographiques sur l'encodage, informations complémentaires pour la compréhension du texte

#### front

Contient tout ce qui est au début du document, avant le corps du texte : page de titre, dédicaces, préfaces, etc.

#### standOff

Élément pour les données liées, les informations contextuelles et les annotations isolées

#### = le paratexte

## Index de personnes

### Le particDesc

Liste de locuteurs, voix ou autres participants identifiables d'une interaction linguistique. Possibilité de compléter avec de **nombreuses informations** (relations, lieux de vie, âge, etc).

#### Le listRelation

```
<listPerson>
   <person xml:id="husband" sex="1">
      <persName>Monsieur</persName>
   </person>
   <person xml:id="wife" sex="2">
       <persName>Madame</persName>
   </person>
   <listRelation>
       <relation type="spouse" mutual="#husband #wife"/>
    </listRelation>
</listPerson>
```

## Exercice

Créer un index de personnages

## Consigne

En vous appuyant sur les <u>recommandations</u> et les <u>exemples</u> fournis dans les <u>guidelines</u>, créer un index de personnage pour l'extrait de <u>Notre Dame de Paris</u> en balisant les mentions et interventions des différents locuteurs avec leurs identifiants

Gringoire Monsieur le prévôt Maître Guillaume Doux-Sire

## Index de lieux

## Le settingDesc

Liste de lieux dans lesquels se situe l'intrigue, soit sous forme d'une description en prose, soit d'une série d'éléments. Possibilité de compléter avec de <u>nombreuses informations</u> (pays, coordonnées, etc).

## Exercice

Créer un index de toponymes

## Consigne

En vous appuyant sur les <u>recommandations</u> et les <u>exemples</u> fournis dans les *guidelines*, créer un index de toponymes pour l'extrait de <u>Notre Dame de Paris</u> en balisant les mentions des différents lieux avec leurs identifiants

### Visualiser son index

Appliquer un scénario de transformation générique à son encodage

### Transformation avec XSL



- 1. Créer un scénario de transformation
  - 2. Sélectionner "XML transformation with XSLT"
  - 3. Sélectionner dans "XSL URL" le document .xsl de transformation
  - 4. Dans l'onglet "Sortie", préciser un nom de fichier pour le résultat de la transformation
  - 5. Appliquer le scénario
  - 6. Ouvrir le fichier .html avec son navigateur

## **Feuille XSL**

La standardisation de l'encodage induite par le respect des *guidelines* permet d'utiliser des <u>scénarios de transformation génériques</u>