Séance 5

Les TEI Guidelines

Modules, modèles, macros, etc.

Correction

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

Principes des guidelines

Dans quel but et pour quel usage ont été constituées les recommandations

Les principes de Poughkeepsie

- Fournir un format standard
- 2. Favoriser l'**échange de textes** dans les humanités
- Suggérer des principes abstraits pour l'encodage des textes
- 4. Inclure un **ensemble minimal de conventions** pour l'encodage de nouveaux textes
- 5. Proposer des ensembles de conventions d'encodage adaptés à **plusieurs applications différentes**

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

Comment lire les Guidelines?

Exemples

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

Type de texte

En fonction du contenu à encoder, par exemple dans la rubrique "Verse"

Table des matières

En parcourant les différentes rubriques qui détaillent des points précis

Élément

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

Implémentation

Un modèle pour de nombreuses applications :

Modèle abstrait

Ensemble des concepts et les relations entre ces concepts

Implémentation

Retranscription des contraintes formelles par le biais d'un schéma de validation

Structure des guidelines

Comment sont construites les recommandations

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

header

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

gaiji Character and Glyph <u>5 Characters, Glyphs, and Writing Modes</u>

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>11 Representation of Primary Sources</u>

verse Verse <u>6 Verse</u>

Les classes

Les classes permettent d'organiser les éléments du modèle. 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

deux

@corresp,@next,

@who

@prev, @sameAs, etc.

att.ascribed pour discours action pouvant être

attribué à un individu spécifique

Attributs d'une classe

att.written

Home

B Attribute Classes

Module	tei — The TEI Infrastructure				
Members	att.damaged [damage damageSpan] att.textCritical [lem rdg rdgGrp] att.transcriptional [add addSpan del delSpan mod redo restore retrace resubst 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> eleconcerned.	ement describing the hand considered responsible for the content of the element		
		Status	Optional		
		Datatype	teidata.pointer		

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	os 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	lem rdg		
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 <u>macros</u>

Les macros définissent des sortes de "raccourcis" pour remplacer des structures fréquemment utilisées.

Elles sont utilisées de deux façons :

- pour représenter des modèles de contenu ou des parties de modèles de contenu fréquemment rencontrés;
- pour représenter des types de données d'attributs.

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 o 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 writing		
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)*		

Consulter les guidelines

Comment se retrouver sur ce site pas très engageant

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 Colomban

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
- 16 Linking, Segmentation, and Alignment
- **18 Feature Structures**
- 19 Graphs, Networks, and Trees
- 20 Non-hierarchical Structures
- **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. 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 Deprecations
- Appendix H <u>Prefatory Notes</u>

Appendix I Colophon

Text Body

- 3 Elements Available in All TEI Documents
- 4 Default Text Structure
- S Characters, Glyphs, and Writing Modes
- ⊕ 6 Verse
- 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
- **17 Simple Analytic Mechanisms**
- **18 Feature Structures**
- 19 Graphs, Networks, and Trees
- 20 <u>Non-hierarchical Structures</u>
- 21 <u>Certainty, Precision, and Responsibility</u>
- **22 Documentation Elements**

TEI sourcecode

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

Documentation des modules

Annexes

Home C Elements

Module	textstructure — Default Text Structure		
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 (@cha (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		

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	att.global (@xml:id, @ attribute class				
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 &
```

Modèle de contenu

```
<content>
  <alternate min0ccurs="0"
    max0ccurs="unbounded">
    <textNode/>
    <classRef key="model.gLike"/>
    <classRef key="model.graphicLike"/>
    <classRef key="model.global"/>
    <elementRef key="surface"/>
    <classRef key="model.linePart"/>
    </alternate>
  </content>
```

Définition schematron du modèle de contenu de l'élément zone

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 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

Consigne

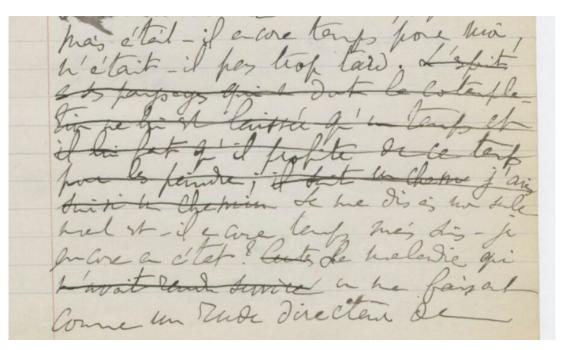
En naviguant dans les *Guidelines* par les différents moyens évoqués, trouver trois différentes balises pouvant être utilisées pour encoder une prise de parole

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>



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

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

Exercice

Créer une bibliographie

Consigne

Choisissez trois ressources 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

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

Le publicationStmt

Méthode conditionnée par où l'on place notre document dans la chaîne éditoriale

Référence bibliographique non-structurée

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

Méthode la plus simple

bibl

```
<sourceDesc>
    <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>
</sourceDesc>
```

biblFull

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>

<biblFull>...</biblFull>

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

- 1. Gringoire
- 2. Monsieur le prévôt
- 3. Maître Guillaume Doux-Sire

Pour aller plus loin sur l'encodage des <u>noms de personnes</u>

Le particDesc

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

Le front matter

Liste de personnages, particulièrement adapté lorsque le texte édité présente en préambule les protagonistes

Le standOff

Pour isoler certaines informations contextuelles. À placer directement dans l'élément TEI :

TEI/standOff

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"/>
    </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

Index de lieux

- 3. Port-au-foin
- 4. Paris
- 5. Sainte-Chapelle
- 6. Rue de la Savaterie
- 7. Île de la Cité

- Palais de Justice 1. Rue de la Barillerie
- 2. Rue Grenier-sur-l'Eau 2. Rue de la Vieille Draperie
 - 3. Rue de la Juiverie
 - 4. Pont Saint-Michel
 - 5. Pont au Change
 - 6. Place de Grève
 - 7. Seine

Pour aller plus loin sur l'encodage des noms de lieux

Le settingDesc

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

Le front matter

Liste de lieux, particulièrement adapté lorsque le texte édité présente en préambule les endroits où se déroule l'intrigue

Le standOff

Pour isoler certaines informations contextuelles. À placer directement dans l'élément TEI :

TEI/standOff

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

Feuille XSL

La standardisation de l'encodage induite par le respect des guidelines permet d'utiliser des scénarios de transformation génériques

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