

# **Introduction à XML-TEI**

## **Séance 1**

**RCL - UVSQ - 2024-25**

# **Pour le parcours Recherche et création littéraire (M1 - M2)**

**Mercredi 10h-12h, A302**

## **Séance 1**

Présentation de XML-TEI

## **Séance 2**

Encodage d'un article de revue, d'une poésie, d'une pièce de théâtre. Présentation des attributs

## **Séance 3**

Pratique sur des textes de la Renaissance

## **Séance 4**

Contraindre le XML : présentation de la DTD

## **Séance 5**

Présentation du TeiHeader

## **Séance 6**

Présentation de l'ODD

## **Séance 7**

De l'HTR à l'édition XML-TEI

## **Séance 8**

Évaluation



# À quoi sert la TEI ?

## Text Encoding Initiative

Pourquoi la TEI est-elle utile?

- Préservation et accessibilité
- Interopérabilité
- Analyse textuelle avancée
- Documentation et annotation

# Exemples d'un encodage XML-TEI :

- Arrêt du Parlement de Bordeaux, 17 décembre 1562
- L'année 1437 dans la pratique de Pierre Christofle, notaire du Châtelet d'Orléans :
  - la source
  - l'encodage XML-TEI
  - le résultat



# Le texte informatisé

# Texte informatisé

## Texte brut

Bonjour à tous. Aujourd'hui, nous allons étudier les bases de la TEI.



# Texte informatisé

## Texte balisé

```
<document>
```

```
  <title>Introduction à la TEI</title>
```

```
  <body>
```

```
    <p>Bonjour à tous. <emphasis> Aujourd'hui, nous allons étudier</emphasis>  
    les bases de la TEI.</p>
```

```
  </body>
```

```
</document>
```

# Le texte informatisé

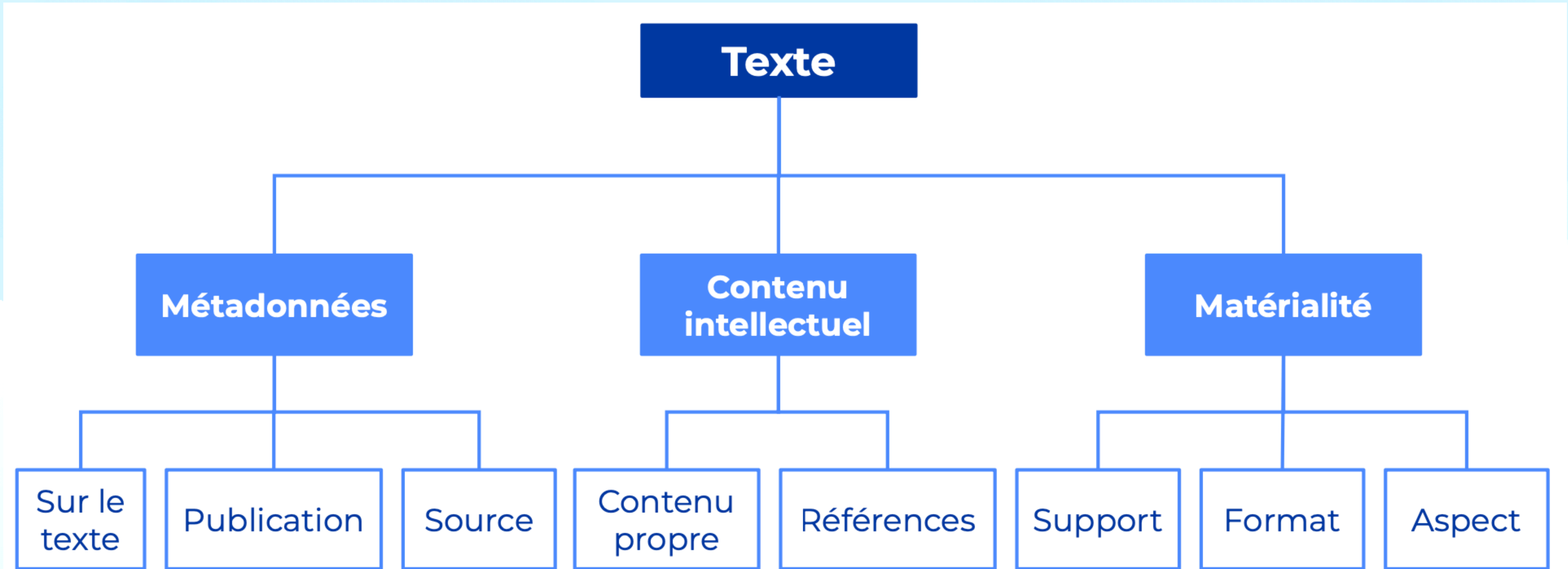
## Texte formaté

```
<!DOCTYPE html><html lang="fr">
<head>
  <meta charset="UTF-8">
  <title>Introduction à la TEI</title>
  <style>
    .emphasis {font-style: italic; color: blue; }
  </style>
</head>
<body>
  <h1>Introduction à la TEI</h1>
  <p>Bonjour à tous. <span class="emphasis">Aujourd'hui, nous allons étudier</span> les bases de la TEI.</p>
</body>
</html>
```



**Qu'est-ce qu'un texte?**

# Qu'est-ce qu'un texte?





# Qu'est-ce qu'un texte?

## 1. Métadonnées

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Le Cid</title>
      <author>Pierre Corneille</author>
    </titleStmt>
    <publicationStmt>
      <publisher>Gallimard</publisher>
      <pubPlace>Paris</pubPlace>
      <date>1637</date>
    </publicationStmt>
    <sourceDesc>
      <bibl>Corneille, Pierre. *Le Cid*. Paris: Gallimard, 1637.
    </bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

# Qu'est-ce qu'un texte?

## 2. Contenu intellectuel

```
<text>
  <body>
    <div type="act" n="1">
      <head>Acte I</head>
      <div type="scene" n="1">
        <head>Scène 1</head>
        <sp who="#Rodrigue">
          <p>Ô rage ! ô désespoir ! ô vieillesse ennemie !</p>
        </sp>
      </div>
    </div>
  </body>
</text>
```



# Qu'est-ce qu'un texte?

## 3. Matérialité

```
<surface>
  <graphic url="manuscript_page.jpg" n="1"/>
  <zone ulx="50" uly="100" lrx="250" lry="200">
    <p>ô rage ! ô désespoir ! ô vieillesse ennemie !</p>
  </zone>
</surface>
```

# Caractéristiques du XML-TEI



# Caractéristiques du XML

## Un standard international

Depuis 1998, XML est un langage libre et documenté. C'est un standard respectant les recommandations du W3C, il facilite :

- la lisibilité par les machines ou par l'œil humain ;
- l'échange de données ;
- la migration vers d'autres plateformes, d'autres logiciels, d'autres formats.

Il existe plusieurs langages XML :

- TEI: encodage des documents patrimoniaux
- EAD: description des fonds archivistiques
- XHTML: structure d'une page web
- SVG: documents vectoriels

# Caractéristiques du XML

## Un langage à balise

Haïku  
Takahama Kyoshi

Le serpent s'esquiva  
Mais le regard qu'il me lança  
Resta dans l'herbe.

```
{poeme}  
  {meta}  
{type}Haïku{type}  
{auteur}Takahama Kyoshi{auteur}  
  {meta}  
  {contenu}  
{vers}Le serpent s'esquiva{vers}  
{vers}Mais {gras}le regard{gras} qu'il me lança{vers}  
{vers}Resta dans l'herbe.{vers}  
  {contenu}  
{poeme}
```



# Principes fondamentaux du XML

## 1. La déclaration XML

```
<?xml version="1.0" encoding="UTF-8"?>
```

# Principes fondamentaux du XML

## 2. Les balises

```
<p>Bonjour, ceci est un paragraphe.</p>
```

```
<title>Le Cid</title>
```



# Principes fondamentaux du XML

## 3. Imbrication des balises

Règle d'imbrication :

```
<parent>
  <enfant>
    <petit-enfant></petit-enfant>
  </enfant>
</parent>
```

Exemple en TEI:

```
<text>
  <body>
    <div type="act" n="1">
      <head>Acte I</head>
      <p>Voici le début de l'acte I.</p>
    </div>
  </body>
</text>
```

Exemple correct :

```
<p>Ce texte est <hi rend="italic">important</hi>.</p>
```

Exemple incorrect :

```
<p>Ce texte est <hi rend="italic">important.</p></hi>
```

# Principes fondamentaux du XML

## 4. Héritage des styles

```
<text xml:lang="fr">  
  <body>  
    <p>Voici un paragraphe en français.</p>  
  </body>  
</text>
```



# Principes fondamentaux du XML

## 5. Les attributs

```
<p xml:id="p1" rend="italic">Ceci est un paragraphe en  
italique.</p>
```

# Principes fondamentaux du XML

## 6. Auto-fermantes et balises vides

```
<lb n="1"/>  
<lb n="2"/>
```



# Principes fondamentaux du XML

## 7. Importance de la sémantique

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Le Cid</title>
      <author>Pierre Corneille</author>
    </titleStmt>
  </fileDesc>
</teiHeader>
```

# Exercice

À l'aide des balises suivantes, structurer le texte proposé

- `<texte>`
- `<chapitre>`
- `<titreChapitre>`
- `<paragraphe>`
- `<nomPersonnage>`
- `<noteDeBasDePage>`

## **Blanche-Neige mange la pomme empoisonnée**

Il était une fois Blanche-Neige<sup>1</sup> que sa belle-mère détestait. [...]

## **Blanche-Neige attend le prince Charmant**

Un jour Charmant<sup>2</sup> arriva sur son cheval blanc et la sauva. [...]

1 Elle porte ce nom, car son visage est blanc comme la neige

2 Nom du prince qui sauve Blanche-Neige.



# Exercice

## En version XML-TEI

```
<?xml version="1.0" encoding="UTF-8"?>
<texte>
  <chapitre>
    <titreChapitre>Blanche-Neige mange la pomme empoisonnée</titreChapitre>
    <paragraphe>Il était une fois <nomPersonnage>Blanche-Neige</nomPersonnage>
      <noteDeBasDePage>Elle porte ce nom, car son visage est blanc comme la
        neige</noteDeBasDePage> que sa belle-mère détestait </paragraphe>
  </chapitre>
  <chapitre>
    <titreChapitre>Blanche-Neige attend le nain Charmant</titreChapitre>
    <paragraphe>Un jour, <nomPersonnage>Charmant</nomPersonnage><noteDeBasDePage>Nom du prince
      qui sauve Blanche-Neige</noteDeBasDePage>arriva sur son cheval blanc et la sauva.
  </paragraphe>
</chapitre>
</texte>
```

# Les composants du XML-TEI



# Les composants du XML-TEI

## 1. Éléments

### a. Éléments de Structure

```
<text>
  <body>
    <div type="chapter" n="1">
      <head>Chapitre 1</head>
      <p>Ceci est le texte du premier chapitre.</p>
    </div>
  </body>
</text>
```

### b. Éléments de Contenu

```
<person>
  <name>Jean Valjean</name>
  <role>Protagoniste</role>
</person>
```

# Les composants du XML-TEI

## 2. Attributs

### a. Attributs pour la Mise en Forme

```
<p rend="italic">Ceci est un texte en italique.</p>
```

### b. Attributs pour la Classification

```
<person xml:id="p1" role="author">  
  <name>Victor Hugo</name>  
</person>
```



# Les composants du XML-TEI

## 3. Entités

```
<p>Le symbole d&rsquo;apostrophe est &rsquo; dans le texte.</p>
```

```
<!ENTITY copy "©">
```

```
<p>Les droits d&rsquo;auteur sont protégés par &copy; 2024.</p>
```

# Les composants du XML-TEI

## 4. Commentaire

### a. Commentaires Simples

```
<!-- Ceci est un commentaire qui explique la section suivante -->
<text>
  <body>
    <p>Ceci est un paragraphe du texte.</p>
  </body>
</text>
```

### b. Commentaires pour Documentation

```
<text>
  <!-- Ce bloc contient le texte principal du document -->
  <body>
    <div type="section">
      <head>Introduction</head>
      <p>Le texte de l'introduction commence ici.</p>
    </div>
  </body>
</text>
```



# Exercices

## Des encodages bien formés?

Un encodage bien formé suit les règles suivantes :

- 1 balise ouvrante = 1 balise fermante
- respect de l'imbrication
- un attribut du même nom par balise
- un seul élément racine

# Exercices

## Des encodages bien formés?

```
<paragraphe>du texte</paragraphe>  
<paragraphe><article>du</article><nom>texte</nom></paragraphe>  
<paragraphe><article>du <nom></article>texte</nom></paragraphe>  
<paragraphe><article>du <nom>texte</nom></article></paragraphe>  
<paragraphe type="texte">du texte</paragraphe>  
<paragraphe type=texte>du texte</paragraphe>  
<paragraphe type="texte">du texte<paragraphe/>  
<paragraphe type="texte">du texte<nom>nom</paragraphe>  
<paragraphe type="texte">du texte</Paragraphe>  
<segment type="texte" type="nombre">du texte</paragraphe>
```



# Exercices

## 1. Utilisation des éléments

### Éléments à utiliser

- `<text>`
- `<chapitre>`
- `<titreChapitre>`
- `<paragraphe>`
- `<nomPersonnage>`

### Le Petit Poucet trouve les bottes de sept lieues

Il était une fois Le Petit Poucet que ses parents abandonnèrent dans la forêt. [...]

### Le Petit Poucet trompe l'ogre

Un jour, Poucet mit les bottes de l'ogre et sauva ses frères. [...]

# Exercices

## 1. Utilisation des éléments

```
<?xml version="1.0" encoding="UTF-8"?>
<texte>
  <chapitre>
    <titreChapitre>Le Petit Poucet trouve les bottes de sept lieues</
titreChapitre>
    <paragraphe>Il était une fois <nomPersonnage>Le Petit Poucet</
nomPersonnage>
      que ses parents abandonnèrent dans la forêt.</paragraphe>
    </chapitre>
    <chapitre>
      <titreChapitre>Le Petit Poucet trompe l'ogre</titreChapitre>
      <paragraphe>Un jour, <nomPersonnage>Poucet</nomPersonnage> mit les bottes
de l'ogre et sauva ses frères.</paragraphe>
    </chapitre>
  </texte>
```



# Exercices

## 2. Ajout d'attributs

### Éléments et attributs à utiliser

- `<text>`
- `<chapitre>`
- `<titreChapitre>`
- `<paragraphe>`
- `<nomPersonnage>`
- `<noteDeBasDePage>`
- `@type`, pour qualifier les personnage
- `@id`, pour indiquer le numéro de la note

### Cendrillon perd sa pantoufle de verre

Il était une fois Cendrillon<sup>3</sup> qui vivait avec ses horribles demi-sœurs. [...]

### Cendrillon retrouve son prince

Le prince<sup>4</sup> trouva la pantoufle et chercha la jeune fille dans tout le royaume. [...]

3. Cendrillon vivait avec ses horribles demi-sœurs après la mort de son père.

4. Le prince, charmé par Cendrillon, fit tout pour la retrouver.

# Exercices

## 2. Ajout d'attributs

```
<?xml version="1.0" encoding="UTF-8"?>
<texte>
  <chapitre>
    <titreChapitre>Cendrillon perd sa pantoufle de verre</titreChapitre>
    <paragraphe>Il était une fois <nomPersonnage type="protagoniste">Cendrillon</nomPersonnage>
      <noteDeBasDePage id="3">Cendrillon vivait avec ses horribles demi-sœurs après la mort de son
père.</noteDeBasDePage>
      qui vivait avec ses horribles demi-sœurs.</paragraphe>
    </chapitre>
    <chapitre>
      <titreChapitre>Cendrillon retrouve son prince</titreChapitre>
      <paragraphe>Le <nomPersonnage type="antagoniste">prince</nomPersonnage>
        <noteDeBasDePage id="4">Le prince, charmé par Cendrillon, fit tout pour la retrouver.</
noteDeBasDePage>
        trouva la pantoufle et chercha la jeune fille dans tout le royaume.</paragraphe>
      </chapitre>
    </texte>
```



# Exercices

## 3. Inclusion de commentaires

### Éléments et attributs à utiliser

- `<text>`
- `<chapitre>`
- `<titreChapitre>`
- `<paragraphe>`
- commenter la place du paragraphe dans le conte

### La Belle au Bois Dormant reçoit une malédiction

Il était une fois une princesse qui reçut une malédiction d'une fée jalouse. [...]

### La Belle au Bois Dormant se réveille

Un prince la réveilla avec un baiser d'amour sincère. [...]

# Exercices

## 3. Inclusion de commentaires

```
<?xml version="1.0" encoding="UTF-8"?>
<texte>
  <chapitre>
    <titreChapitre>La Belle au Bois Dormant reçoit une malédiction</titreChapitre>
    <!-- Ce paragraphe décrit le début du conte -->
    <paragraphe>Il était une fois une princesse qui reçut une malédiction d'une fée jalouse.</paragraphe>
  </chapitre>
  <chapitre>
    <titreChapitre>La Belle au Bois Dormant se réveille</titreChapitre>
    <!-- Ce paragraphe décrit la conclusion du conte -->
    <paragraphe>Un prince la réveilla avec un baiser d'amour sincère.</paragraphe>
  </chapitre>
</texte>
```



# Exercices

## 4. Vers un texte plus enrichi

### Éléments et attributs à utiliser

- `<texte>`
- `<chapitre>`
- `<titreChapitre>`
- `<paragraphe>`
- commenter la place du paragraphe dans le déroulement de l'histoire
- `<nomPersonnage>` avec attribut `@id` pour les références
- `<noteDeBasDePage>` avec attribut `@id` pour les notes explicatives
- Mettre **quête** et **périls** en gras avec `<rend="bold">`

### Le Chevalier de la Lune reçoit une quête

Dans une nuit sans étoiles, le Chevalier de la Lune<sup>1</sup> fut convoqué par le roi pour accomplir une quête sacrée.

### Le Chevalier de la Lune affronte le Serpent des Abysses

Après de nombreux périls, le chevalier se retrouva face au Serpent des Abysses<sup>2</sup>, gardien de l'entrée du royaume perdu.

1. Le Chevalier de la Lune est un personnage mythique, symbolisant le lien entre la lumière et l'ombre.

2. Le Serpent des Abysses est une créature ancestrale, symbole des dangers cachés dans les profondeurs de l'âme humaine.

# Exercices

## 4. Vers un texte plus enrichi

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<texte>
```

```
  <chapitre>
```

```
    <titreChapitre>Le Chevalier de la Lune reçoit une <hi rend="bold">quête</hi></titreChapitre>
```

```
    <!-- Ce paragraphe présente le point de départ de l'aventure, où le héros accepte son destin -->
```

```
      <paragraphe>Dans une nuit sans étoiles, <nomPersonnage id="1">le Chevalier de la Lune </nomPersonnage><noteDeBasDePage id="1"><nomPersonnage id="1">Le Chevalier de la Lune </nomPersonnage>est un personnage mythique, symbolisant le lien entre la lumière et l'ombre.</noteDeBasDePage> fut convoqué par le roi pour accomplir une <hi rend="bold">quête</hi> sacrée.</paragraphe>
```

```
    </chapitre>
```

```
  <chapitre>
```

```
    <titreChapitre>Le Chevalier de la Lune affronte le Serpent des Abysses</titreChapitre>
```

```
    <!-- Ce paragraphe décrit l'apogée du conte, où le héros doit faire face à son plus grand adversaire -->
```

```
      <paragraphe>Après de nombreux <hi rend="bold">périls</hi>, <nomPersonnage id="1">le Chevalier de la Lune</nomPersonnage> se retrouva face au <nomPersonnage id="2">Serpent des Abysses</nomPersonnage><noteDeBasDePage id="2"><nomPersonnage id="2">Le Serpent des Abysses</nomPersonnage> est une créature ancestrale, symbole des dangers cachés dans les profondeurs de l'âme humaine.</noteDeBasDePage>, gardien de l'entrée du royaume perdu.</paragraphe>
```

```
    </chapitre>
```

```
</texte>
```