

Variété des structures textuelles

Un encodage bien formé?

Retour sur le cours précédent avec un petit exercice.

Pour rappel, 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

Exercice

Bien formé ou mal formé ?

```
1.
<livre>
  <titre>XML pour les débutants</titre>
  <auteur>Jean Dupont</auteur>
  <prix devise=EUR>29.99</prix>
</livre>

2.
<livre>
  <titre>Introduction à XML</titre>
  <auteur><nom><prenom>Albert</prenom>Einstein<nom></auteur>
  <prix devise="USD">39.99</prix>
</livre>

3.
<livre>
  <titre>Introduction à XML</titre>
  <auteur type="célèbre" type="allemand">Albert Einstein</auteur>
</livre>

4.
<commande>
  <produit id="106">
    <nom>Livre</nom>
    <quantite>1</quantite>
    <prix devise="EUR">12.99</prix>
  </produit>
  <produit id="107">
    <nom>Livre</nom>
    <quantite>3</quantite>
    <prix devise="EUR">89.99</prix>
  </produit>
</commande>
```

Variété des structures textuelles

Pour la TEI, les textes et leurs divisions peuvent contenir un éventail assez limité de composants « structurels ».

Ceux-ci incluent les en-têtes et matériaux préliminaires au début ou à la fin d'une division, et un certain nombre de composants de base qui sont caractéristiques de la prose, des vers ou des textes dramatiques. La prose, pour l'essentiel, consiste en des paragraphes ou des listes, marqués en utilisant les éléments `<p>` ou `<list>` respectivement. Les vers consistent en des lignes de vers, marquées à l'aide de l'élément `<l>`, ou des séquences de lignes de vers marquées à l'aide de l'élément `<lg>` (pour « line group », ou groupe de lignes). Dans les textes dramatiques, un composant additionnel est offert sous la forme de l'élément `<sp>` (« speech »), qui combine un élément `<speaker>` (locuteur) avec un ou plusieurs éléments `<p>` ou `<l>`, selon que la pièce est en vers ou en prose.

Encoder un article scientifique

Rappel : La prose, pour l'essentiel, consiste en des paragraphes ou des listes, marqués en utilisant les éléments `<p>` ou `<list>` respectivement. Les éléments d'une liste sont listés par `<item>`. Les titres sont indiqués par `<head>` et les paragraphes par `<div>`.

Exercice

Encoder le texte suivant :

Introduction

Nous recommandons l'utilisation du balisage TEI pour représenter la prose scientifique. Il présente un certain nombre d'avantages pratiques :

- Le balisage TEI est facile à ajouter ;
- Le balisage TEI est largement compris ;
- Le balisage TEI est facile à convertir en d'autres formats.

Le balisage TEI présente aussi certains avantages scientifiques, que nous discuterons dans une autre section.

Origines de la Tei

La Text Encoding Initiative est née en 1987...

Propriétés scientifiques du balisage TEI

Le balisage TEI exprime une vision de la nature du texte...

En XML-TEI :

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <div>
      <head>Introduction</head>
      <p>Nous recommandons l'utilisation du balisage TEI pour
représenter la prose scientifique. Il présente un certain nombre
d'avantages pratiques :</p>
      <list>
```

```

        <item>Le balisage TEI est facile à ajouter ;</item>
        <item>Le balisage TEI est largement compris ;</item>
        <item>Le balisage TEI est facile à convertir en d'autres
formats.</item>
    </list>
    <p>Le balisage TEI présente aussi certains avantages
scientifiques, que nous discuterons dans une autre section.</p>
    <!-- ici les paragraphes d'introduction suivants -->
</div>
<div>
    <head>Origines de la TEI</head>
    <p>La Text Encoding Initiative est née en 1987...</p>
    <!-- ici plein d'autres paragraphes -->
</div>
<div>
    <head>Propriétés scientifiques du balisage TEI</head>
    <p>Le balisage TEI exprime une vision de la nature du texte...
</p>
    <!-- ici plein d'autres paragraphes -->
</div>
</body>
</text>

```

Notez que cet encodage indique seulement l'organisation du document. Il ne dit rien de la façon dont il doit être visualisé sur écran ou sur papier. Il indique qu'il y a des sections qui ont des titres et qui contiennent des paragraphes et des listes. Il distingue les items dans la liste, mais il ne précise pas s'ils doivent être précédés d'une puce, d'un tiret ou d'un numéro.

Comment choisir son encodage? Les Guidelines en TEI

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.

Exemple de l'élément `<div>`

Montrer l'exemple.

Les modules

Les modules sont un ensemble cohérent de balises regroupées pour répondre à une problématique précise. Il y a des modules pour :

- un type de document : modules pour la poésie, le théâtre, les manuscrits, le discours oral, etc
- un type de données : modules pour la définition d'un appareil critique, formules mathématiques, images, etc.

Chaque module :

- est documenté par un chapitre des Guidelines

- définit un certain nombre de composantes
- pointe vers des spécifications techniques
- correspond à une problématique particulière
- peut être inclus ou non dans une personnalisation

Encoder un texte poétique

En naviguant dans les Guidelines par les différents moyens évoqués, trouvez des balises pouvant être utilisées pour encoder un vers et un groupe de vers.

Solution : les vers sont marqués à l'aide l'élément `<l>`, les séquences de lignes de vers par l'élément `<lg>`.

Exercice

Encoder le texte suivant :

Chanson d'automne

Les sanglots longs
Des violons
De l'automne
Blessent mon cœur
D'une langueur
Monotone.

Tout suffocant
Et blême, quand
Sonne l'heure,
Je me souviens
Des jours anciens
Et je pleure.

Et je m'en vais
Au vent mauvais
Qui m'emporte
Deçà, delà,
Pareil à la
Feuille morte.

En XML-TEI :

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <head>Chanson d'automne</head>
    <lg>
      <l>Les sanglots longs</l>
      <l>Des violons</l>
      <l>De l'automne</l>
      <l>Blessent mon cœur</l>
      <l>D'une langueur</l>
```

```

        <l>Monotone.</l>
    </lg>
    <lg>
        <l>Tout suffocant</l>
        <l>Et blême, quand</l>
        <l>Sonne l'heure,</l>
        <l>Je me souviens</l>
        <l>Des jours anciens</l>
        <l>Et je pleure.</l>
    </lg>
    <lg>
        <l>Et je m'en vais</l>
        <l>Au vent mauvais</l>
        <l>Qui m'emporte</l>
        <l>Deçà, delà,</l>
        <l>Pareil à la</l>
        <l>Feuille morte.</l>
    </lg>
</body>
</text>

```

Encoder une pièce de théâtre

Dans les Guidelines, trouvez le meilleur encodage pour cet extrait.

Dans une pièce de théâtre, un composant additionnel est offert sous la forme de l'élément `<sp>` (« speech »), qui combine un élément `<speaker>` (locuteur) avec un ou plusieurs éléments `<p>` ou `<l>`, selon que la pièce est en vers ou en prose. Il existe également l'élément `<stage>` pour des descriptions de scène ou des indications scéniques.

Exercice

Encoder le texte suivant :

Le Cid
 Scène de confrontation entre Don Rodrigue et Chimène.

Don Rodrigue:
 Je suis jeune, il est vrai ; mais aux âmes bien nées,
 La valeur n'attend pas le nombre des années.

Chimène:
 Vous me parlez de moi-même ; et cependant,
 Je vois bien que ce n'est pas vous que vous défendez.

Don Rodrigue:
 Il est vrai que je ne me défends pas moi-même,
 Mais je défends ma cause avec tout mon cœur.

En XML-TEI :

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <head>Le Cid</head>
    <stage>Scène de confrontation entre Don Rodrigue et Chimène.
  </stage>
    <sp>
      <speaker>Don Rodrigue</speaker>
      <p>Je suis jeune, il est vrai ; mais aux âmes bien nées,</p>
      <p>La valeur n'attend pas le nombre des années.</p>
    </sp>
    <sp>
      <speaker>Chimène</speaker>
      <p>Vous me parlez de moi-même ; et cependant,</p>
      <p>Je vois bien que ce n'est pas vous que vous défendez.</p>
    </sp>
    <sp>
      <speaker>Don Rodrigue</speaker>
      <p>Il est vrai que je ne me défends pas moi-même,</p>
      <p>Mais je défends ma cause avec tout mon cœur.</p>
    </sp>
  </body>
</text>
```

Remarque : œ est une entité, elle doit devenir `&oeelig;`.

Encoder une lettre

Dans les Guidelines, trouvez le meilleur encodage pour cet extrait. *La réponse est dans le "4. Default Text Structure", au "4.2.2 Openers and Closers"

Dans une lettre, des composants spécifiques sont proposés pour structurer ses différentes parties :

- `<div type="letter">` (« lettre ») : Ce composant encapsule l'intégralité de la lettre, indiquant qu'il s'agit d'une correspondance.
- `<opener>` (« ouverture ») : Cet élément regroupe les informations en tête de la lettre, généralement utilisées pour introduire le texte.
 - `<dateline>` (« date et lieu ») : Contient la date et le lieu où la lettre a été écrite.
 - `<salute>` (« salutation d'ouverture ») : Représente la formule de salutation initiale adressée au destinataire.
- `<p>` (« paragraphe ») : Encadre chaque paragraphe du corps principal de la lettre.
- `<closer>` (« clôture ») : Ce composant regroupe les éléments qui clôturent la lettre.
 - `<salute>` (« salutation finale ») : Indique la formule de politesse finale.
 - `<signed>` (« signature ») : Contient la signature de l'auteur de la lettre.

Exercice

Encoder le texte suivant :

Paris, le 15 août 2024

Chere Amie,

J'espère que cette lettre te trouve en bonne santé. Je voulais te parler de quelques idées pour notre projet. Voici ce que je propose :

- Nous devons nous concentrer sur l'analyse des données existantes ;
- Il serait judicieux d'envisager une collaboration avec d'autres équipes ;
- Enfin, un budget détaillé est nécessaire avant de procéder.

Je te tiendrai informée des prochaines étapes.

Bien à toi,
Jean Dupont

En XML-TEI :

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <div type="letter">
      <!-- En-tête de la lettre, incluant la date et le lieu -->
      <opener>
        <dateline>Paris, le 15 août 2024</dateline>
        <salute>Cher(e) Ami(e),</salute>
      </opener>

      <!-- Contenu principal de la lettre -->
      <p>J'espère que cette lettre te trouve en bonne santé. Je
voulais te parler de quelques idées pour notre projet. Voici ce que je
propose :</p>
      <list>
        <item>Nous devons nous concentrer sur l'analyse des
données existantes ;</item>
        <item>Il serait judicieux d'envisager une collaboration
avec d'autres équipes ;</item>
        <item>Enfin, un budget détaillé est nécessaire avant de
procéder.</item>
      </list>
      <p>Je te tiendrai informé(e) des prochaines étapes.</p>

      <!-- Clôture de la lettre, incluant les salutations et la
signature -->
      <closer>
        <salute>Bien à toi,</salute>
        <signed>Jean Dupont</signed>
      </closer>
    </div>
  </body>
</text>
```

Enrichir le texte avec les attributs

1. Introduction aux Attributs en XML

Définition des Attributs :

En XML, les attributs fournissent des informations supplémentaires sur un élément. Ils sont utilisés pour ajouter des métadonnées ou des détails spécifiques qui ne sont pas directement représentés par le contenu de l'élément.

Syntaxe :

Les attributs sont ajoutés à une balise d'élément à l'intérieur des guillemets après le nom de l'élément.

Exemple de base :

```
<element attribut="valeur">Contenu</element>
```

2. Premiers exemples avec des Attributs en TEI

a. Exemple 1 : Attribut **type**

Contexte : Utilisé pour spécifier le type d'un élément, comme le type d'un texte ou d'un div.

Exemple TEI :

```
<text>
  <body>
    <div type="poème">
      <p>Voici un poème.</p>
    </div>
  </body>
</text>
```

Explication : L'attribut **type="poème"** précise que le **<div>** contient de la poésie. Cela aide à classer le contenu.

b. Exemple 2 : Attribut **n** pour les Numéros

Contexte : Utilisé pour numéroter des éléments comme des paragraphes ou des sections.

Exemple TEI :

```
<text>
  <body>
    <p n="1">Premier paragraphe.</p>
    <p n="2">Deuxième paragraphe.</p>
  </body>
```



```
</text>
```

Explication : L'attribut `n` est utilisé pour attribuer un numéro unique à chaque paragraphe. Cela peut être utile pour référencer ou organiser le texte.

3. Exemples Progressifs avec des Attributs

a. Exemple de Poésie avec des Attributs

Poème Simple :

```
<text>
  <body>
    <div type="poème">
      <l n="1">Voici un poème.</l>
      <l n="2">Chaque ligne est balisée.</l>
    </div>
  </body>
</text>
```

Explication : Ici, `<l>` est utilisé pour chaque ligne du poème avec un attribut `n` pour numérotter les lignes.

b. Exemple de Pièce de Théâtre avec des Attributs

Pièce Simple :

```
<text>
  <body>
    <div type="play">
      <sp who="#speaker1">
        <speaker>Personnage 1</speaker>
        <p>Bonjour, comment allez-vous ?</p>
      </sp>
      <sp who="#speaker2">
        <speaker>Personnage 2</speaker>
        <p>Je vais bien, merci !</p>
      </sp>
    </div>
  </body>
</text>
```

Explication : L'attribut `who` dans `<sp>` spécifie l'identifiant du locuteur, permettant d'associer chaque ligne de dialogue à un personnage.

4. Attributs pour le Référencement et la Structure

Exemple de Référencement avec des Attributs

Document Structuré :

```
<text>
  <body>
    <div type="section" n="1">
      <head>Introduction</head>
      <p>Le contenu de l'introduction...</p>
    </div>
    <div type="section" n="2">
      <head>Méthodologie</head>
      <p>Le contenu de la méthodologie...</p>
    </div>
  </body>
</text>
```

Explication : L'attribut `n` dans `<div>` est utilisé pour numéroter les sections du document, facilitant la navigation et le référencement.

Exercice : encoder un poème

Consigne : Encoder les éléments et les attributs pour le poème *Mon rêve familial* de Paul Verlaine.

Mon rêve familial

Je fais souvent ce rêve étrange et pénétrant
D'une femme inconnue, et que j'aime, et qui m'aime
Et qui n'est, chaque fois, ni tout à fait la même
Ni tout à fait une autre, et m'aime et me comprend.

Car elle me comprend, et mon coeur, transparent
Pour elle seule, hélas ! cesse d'être un problème
Pour elle seule, et les moiteurs de mon front blême,
Elle seule les sait rafraîchir, en pleurant.

Est-elle brune, blonde ou rousse ? – Je l'ignore.
Son nom ? Je me souviens qu'il est doux et sonore
Comme ceux des aimés que la Vie exila.

Son regard est pareil au regard des statues,
Et, pour sa voix, lointaine, et calme, et grave, elle a
L'inflexion des voix chères qui se sont tues.

Proposition d'encodage :

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

```

<lg type="sonnet">
  <head>Mon rêve familial</head>
  <lg type="quatrain" n="1">
    <l>Je fais souvent ce rêve étrange et pénétrant</l>
    <l>D'une femme inconnue, et que j'aime, et qui m'aime</l>
    <l>Et qui n'est, chaque fois, ni tout à fait la même</l>
    <l>Ni tout à fait une autre, et m'aime et me comprend.</l>
  </lg>
  <lg type="quatrain" n="2">
    <l>Car elle me comprend, et mon coeur, transparent</l>
    <l>Pour elle seule, hélas ! cesse d'être un problème</l>
    <l>Pour elle seule, et les moiteurs de mon front blême,</l>
    <l>Elle seule les sait rafraîchir, en pleurant.</l>
  </lg>
  <lg type="sizain">
    <lg type="tercet" n="1">
      <l>Est-elle brune, blonde ou rousse ? – Je l'ignore.</l>
      <l>Son nom ? Je me souviens qu'il est doux et sonore</l>
      <l>Comme ceux des aimés que la Vie exila.</l>
    </lg>
    <lg type="tercet" n="2">
      <l>Son regard est pareil au regard des statues,</l>
      <l>Et, pour sa voix, lointaine, et calme, et grave, elle
a</l>
      <l>L'inflexion des voix chères qui se sont tues.</l>
    </lg>
  </lg>
</body>
</text>

```

Et maintenant avec les rimes :

```

<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <lg type="sonnet">
      <head>Mon rêve familial</head>
      <lg type="quatrain" n="1" rhyme="ABBA">
        <l rhyme="A">Je fais souvent ce rêve étrange et
pénétrant</l>
        <l rhyme="B">D'une femme inconnue, et que j'aime, et qui
m'aime</l>
        <l rhyme="B">Et qui n'est, chaque fois, ni tout à fait la
même</l>
        <l rhyme="A">Ni tout à fait une autre, et m'aime et me
comprend.</l>
      </lg>
      <lg type="quatrain" n="2" rhyme="ABBA">
        <l rhyme="A">Car elle me comprend, et mon coeur,
transparent</l>
        <l rhyme="B">Pour elle seule, hélas ! cesse d'être un

```

```

problème</l>
    <l rhyme="B">Pour elle seule, et les moiteurs de mon front
blême,</l>
    <l rhyme="A">Elle seule les sait rafraîchir, en pleurant.
</l>
</lg>
<lg type="sizain">
    <lg type="tercet" n="1" rhyme="CCD">
        <l rhyme="C">Est-elle brune, blonde ou rousse ? – Je
l'ignore.</l>
        <l rhyme="C">Son nom ? Je me souviens qu'il est doux et
sonore</l>
        <l rhyme="D">Comme ceux des aimés que la Vie exila.</l>
    </lg>
    <lg type="tercet" n="2" rhyme="EDE">
        <l rhyme="E">Son regard est pareil au regard des
statues,</l>
        <l rhyme="D">Et, pour sa voix, lointaine, et calme, et
grave, elle a</l>
        <l rhyme="E">L'inflexion des voix chères qui se sont
tues.</l>
    </lg>
</lg>
</body>
</text>

```

Un encodage complexe : le dictionnaire

Encoder les deux entrées suivantes :

Amour

- **Nom masculin**
- **Sens 1** : Sentiment d'affection intense et profonde envers une personne. (*littéraire*)
 - Exemple : *Il lui déclara son amour lors d'une soirée étoilée.*
 - Référence : Victor Hugo, *Les Misérables*
- **Sens 2** : Passion ou affection intense envers une idée, une activité.
 - Exemple : *Son amour de la liberté était sans bornes.*

Beauté

- **Nom féminin**
- **Sens 1** : Qualité de quelqu'un ou de quelque chose qui est esthétique ou agréable à regarder. (*courant*)
 - Exemple : *La beauté de ce paysage est époustouflante.*
 - Référence : Alphonse de Lamartine, *Méditations poétiques*
- **Sens 2** : Idée de perfection, associée aux formes, couleurs ou proportions harmonieuses. (*philosophie*)
 - Exemple : *Pour Platon, la beauté réside dans l'harmonie des formes et des idées.*

Encodage proposé :

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <body>
    <div type="dictionary">
      <!-- Première entrée du dictionnaire : "amour" -->
      <entry xml:id="e1">
        <form>
          <orth>amour</orth>
          <gramGrp>
            <pos>nom masculin</pos>
            <gen>masculin</gen>
          </gramGrp>
        </form>
        <sense xml:id="s1">
          <def>Sentiment d'affection intense et profonde envers une
personne.</def>
          <usg type="register"><hi rend="italic">littéraire</hi></usg>
          <example>
            <quote><hi rend="italic">Il lui déclara son amour lors d'une
soirée étoilée.</hi></quote>
            <cit>
              <bibl><hi rend="italic">Victor Hugo, <title>Les
Misérables</title></hi></bibl>
            </cit>
          </example>
        </sense>
        <sense xml:id="s2">
          <def>Passion ou affection intense envers une idée, une
activité.</def>
          <example>
            <quote><hi rend="italic">Son amour de la liberté était sans
bornes.</hi></quote>
          </example>
        </sense>
      </entry>

      <!-- Deuxième entrée du dictionnaire : "beauté" -->
      <entry xml:id="e2">
        <form>
          <orth>beauté</orth>
          <gramGrp>
            <pos>nom féminin</pos>
            <gen>féminin</gen>
          </gramGrp>
        </form>
        <sense xml:id="s3">
          <def>Qualité de quelqu'un ou de quelque chose qui est
esthétique ou agréable à regarder.</def>
          <usg type="register"><hi rend="italic">courant</hi></usg>
          <example>
            <quote><hi rend="italic">La beauté de ce paysage est
```

```
époustouflante.</hi></quote>
    <cit>
        <bibl><hi rend="italic">Alphonse de Lamartine,
<title>Méditations poétiques</title></hi></bibl>
    </cit>
</example>
</sense>
<sense xml:id="s4">
    <def>Idée de perfection, associée aux formes, couleurs ou
proportions harmonieuses.</def>
    <usg type="domain"><hi rend="italic">philosophie</hi></usg>
    <example>
        <quote><hi rend="italic">Pour Platon, la beauté réside dans
l'harmonie des formes et des idées.</hi></quote>
    </example>
</sense>
</entry>
</div>
</body>
</text>
```