



# Séance 10

# Les schémas

# XML

Initiation à l'ODD



# Les schémas XML

Les différents langages

# Qu'est-ce qu'un schéma ?

Document définissant un **ensemble de règles** concernant la structure et le contenu de fichiers XML et pouvant permettre leur **validation**

# DTD

Définit la structure d'un document, les éléments et attributs qui y sont autorisés, et le type de contenu permis

## Création

Permet la création d'éléments, de sous-éléments, d'attributs, d'entités

## Typage

Pas de typage précis du contenu des éléments, pas de gestion des espaces de nom

## Format

Format qui n'est pas du XML

# DTD

```
<!ELEMENT texte (chapitre+)>  
<!ELEMENT chapitre (titre?, paragraphe+)>  
<!ELEMENT titre (#PCDATA)>  
<!ATTLIST chapitre n CDATA #REQUIRED>
```

# XML schema

Recommandation par le W3C permettant de définir la structure et le type de contenu d'un document XML

## Typage

Permet de préciser le type et le nombre des données autorisés

## Séquence

Permet de préciser des séquences d'éléments et leur contenu

## Format

Les fichiers XML schema sont eux mêmes des documents XML

# XML schema

```
<xs:element name="texte" maxOccurs="unbounded">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="titre"
                  type="xs:string"/>
      <xs:element name="chap" type="xs:decimal"
                  minOccurs="0"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

# Relax NG

Langage de description de document XML, basé sur Relax (*REgular LAnguage description for XML*) et TREX (*Tree Regular Expressions for XML*)

## Structure

Supporte les XML namespace et peut s'appuyer sur les modules et macros de la TEI

## Contenu

Permet de typer les données autorisées

Peut créer des contraintes d'enchaînement très précises grâce au langage [schematron](#)

## Format

Langage respectant la syntaxe XML et compatible avec XML schema



# Relax NG

```
<element name="edition">
  <zeroOrMore>
    <element name="texte">
      <element name="titre">
        <text/>
      </element>
      <element name="paragraphe">
        <text/>
      </element>
    </element>
  </zeroOrMore>
</element>
```

# Schematron

Langage intégré à Relax NG, il permet de valider la structure d'un document XML par une **liste d'assertions**.

Il permet de définir des contraintes complexes, mais contrairement aux autres langages de validation, tout ce qui n'est pas spécifié est considéré comme valide.

# XML namespace

Utilisé pour **éviter les conflits de noms d'éléments** quand plusieurs espaces de noms sont utilisés ensemble. Tous les éléments utilisant le *namespace* doivent être contenus par un élément où est spécifié l'URI du *namespace* dans les attributs.

## Sans préfixe

```
<TEI xmlns="ns_URI">
  <teiHeader/>
  <text/>
</table>
```

## Avec préfixe

```
<a:div xmlns:a="URI_a">
  <a:h1/>
  <a:text/>
</a:div>

<b:div xmlns:b="URI_b">
  <b:h1/>
  <b:text/>
</b:div>
```

## Double

```
<root xmlns:a="URI_a"
       xmlns:b="URI_b">
  <a:div>
    <a:h1/>
    <a:text/>
  </a:div>

  <b:div>
    <b:h1/>
    <b:text/>
  </b:div>
</root>
```



# Spécification TEI

Le langage ODD

# One Document Does it all

Format XML TEI qui comporte à la fois un **schéma de validation** et une **documentation** de ce même schéma

Le langage ODD a été développé en même temps que la TEI afin de pouvoir **documenter** et **formaliser** les recommandations

Une spécification ODD est un simple document XML TEI qui utilise le module **tagdocs**

# ODD

Document XML qu'il est possible de transformer en fichier RelaxNG (.rng) à l'aide d'une feuille de transformation XSLT ou du logiciel [Roma](#)

## Documentation

Fichier pour générer un schéma RelaxNG et une documentation

## Contenu

Permet de définir des séquences, des nombres d'occurrences, des types de données. Supporte les espaces de noms

## Format

Intégralement en syntaxe XML TEI, il s'appuie sur le schéma TEI All et sa structure en modules et macros

# Contenu d'une ODD

## En-tête

Header TEI contenant les métadonnées sur le fichier et sur les acteurs à responsables de l'encodage

`teiHeader`

## Corps

Documentation en prose et ponctuée d'exemples des choix d'encodage effectués

`body`

## Spécifications

Règles de validation concernant les objets et leurs contenus retenus pour l'encodage

`schemaSpec`

# Comment documenter ?

Chaque questionnement concernant la manière d'encoder une information doit être documenté : **expliquer** pourquoi le choix a été effectué et **définir** un modèle d'encodage qui respecte ce choix

Le chapitre [Use the TEI](#) comment une ODD devrait être définie



# Quoi documenter ?

## Utilisation des balises

Rappeler la structuration d'un document TEI et à quoi servent ses composants, détailler comment les balises ont été utilisées dans l'encodage

## Choix de méthode d'encodage

Expliciter quels choix d'encodage ont été privilégiés et spécifier comment ont été balisés les cas un peu limites ou ambigus

## Objectif visé par l'encodage

Expliquer comment les choix d'encodage ont été motivés par un certain nombre d'objectifs scientifiques ou de projets de valorisation

# Cas d'usage

```
o<choice>  
  <orig>u</orig>  
  <reg>ù</reg>  
</choice> il va
```

## **Expliciter un choix quant à certaines règles suivies lors de l'encodage**

Exemple : Rétablissement des accents pour lever l'ambiguïté entre les mots grammaticaux, notamment entre le "a" et le "à" ou encore le "ou" et le "où".

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom square is a dark blue, and the top square is a lighter blue, positioned slightly to the right and above the bottom one.

# Customisation de la TEI

Modifications possibles des recommandations

# Utiliser la TEI dans un projet

Pour des détails concernant la mise en œuvre de la TEI dans le cadre d'un projet d'encodage spécifique, les Guidelines disposent d'un [chapitre complet](#) sur les customisations, la conformance et sur la création d'ODD

# Livrables pour le devoir

- TEI : contenu de votre édition
- ODD : règles de validation + documentation
- DTD : déclarations d'entités
- RNG : schéma produit avec l'ODD

# Conformance à la TEI

## Fichier XML bien formé

L'encodage doit constituer un fichier XML valide

## Validation par la TEI all

L'encodage doit pouvoir être validé avec un schéma TEI all

## Respect des principes généraux

L'encodage doit respecter le modèle abstrait de la TEI

# Conformance à la TEI

## Fichier XML bien formé

L'encodage doit constituer un fichier XML valide

## Validation par la TEI all

L'encodage doit pouvoir être validé avec un schéma TEI all

## Respect des principes généraux

L'encodage doit respecter le modèle abstrait de la TEI

## Documentation

Les choix d'encodage doivent être documenté dans un document dédié, notamment pour y signaler les écarts par rapport à la TEI all

# Pourquoi customiser la TEI ?

## Uniformiser

Obliger tous les encodages à répondre au même ensemble de règles définies

## Clarifier

Éviter les confusions et l'usage accidentel d'élément dont vous n'avez pas besoin

## Augmenter

Étendre le schéma de la TEI pour répondre à des cas non prévus dans les *Guidelines*

## Restreindre

Définir de manière plus strictes les possibilités d'encodage de cas particuliers



A decorative graphic on the left side of the slide consisting of two overlapping squares. The top square is a lighter blue, and the bottom square is a darker blue.

# Quelques customisations de la TEI

**teilight** TEI allégée

**tei\_corpus** TEI pour les corpus linguistiques

**tei\_ms** TEI pour la description de manuscrits

**tei\_drama** TEI pour les corpus dramatique

**tei\_allPlus** Toute la TEI avec des additions

# Customisations

**Suppression** d'éléments, d'attributs ou de classes entières

**Ajout d'éléments** au sein d'une classe ou changement de classe d'un élément

**Modification du modèle de contenu** d'un élément

	Classe	Élément	Attribut
--	--------	---------	----------

**Ajout d'attributs** ou de valeurs d'attribut admis par une classe ou un élément

Ajout		Création d'un élément, ajout d'un attribut	
-------	--	--	--

**Renommage** d'éléments, modification de leur attribut par rapport aux *Guidelines*

Suppression			
-------------	--	--	--

Modification

Usage, définition, type de contenu, nom



# Types de modification à la TEI

## ***Clean***

Le schéma est un  
sous-ensemble de la  
TEI All (*TEI conformant*)

## ***Unclean***

Le schéma considère  
comme valides des  
documents que la TEI  
All considère invalides

# Cas d'usage

```
<note type="commentaire">  
  Commentaire intéressant  
</note>
```

## **Restreindre les valeurs d'attribut autorisées**

Ex : L'attribut @type de l'élément note admet n'importe quelle chaîne de caractère comme type de contenu, mais on peut vouloir réduire ces valeurs à un petit nombre de possibilités

# Spécification de schéma

Un schéma est défini avec l'élément **<schemaSpec>** contenant un mélange de déclaration d'objets (éléments, classes, macros) et de références à d'autres à d'autres objets contenant ce genre de déclaration

```
(moduleRef | elementSpec | macroSpec | classSpec)*
```

# Définition d'un schéma

```
<schemaSpec ident="schemaName">  
  <!-- liste des modules utilisés -->  
  <moduleRef key="modName"/>  
  <!-- liste des éléments modifiés -->  
  <elementSpec ident="elName" mode="mode"/>  
  <!-- liste des classes modifiées -->  
  <classSpec ident="clsName" mode="mode"/>  
  <!-- liste des macros modifiées -->  
  <macroSpec ident="macName" mode="mode"/>  
</schemaSpec>
```

# Modes de modifications

## Add

Ajouter un objet qui n'existe pas dans la TEI

## Replace

L'objet remplace l'objet préexistant dont le nom correspondant à @ident

## Delete

Toute référence à l'objet dont le nom correspond à @ident

## Change

Les éléments enfants à l'objet sont régis par les nouvelles spécifications

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

# Génération d'ODD

Les méthodes pour automatiser  
la création d'une ODD



# ODD by example

*ODD by example* est un scénario de transformation XSLT fourni par la *Text Encoding Initiative* qui permet de générer automatiquement un fichier ODD à partir d'un fichier encodé TEI

[Tutoriel en français](#)

# Add-on Oxygen

## Mise à jour des add-on dans Oxygen

Options > Préférences > Association de type de document

- Tout désactiver
- Vérifier que la ligne "options globales" est cochée
- Valider avec OK

## Ajout de l'add-on TEI P5

Aide > Installer les nouveaux add-ons

- Ajouter <https://www.tei-c.org/release/oxygen/updateSite.oxygen>
- Appliquer et accepter

Aide > Gérer les Add-ons > Installer

- Tout activer et tout appliquer

Options > Préférences > Association de type de document

- Tout activer, appliquer, accepter

# Mise en place du scénario

## Configurer un scénario de transformation

Document > Transformation > Configurer (Ctrl + Maj + C)

- Créer un nouveau scénario > XML transformation with XSL

- Renseigner le chemin XSL

`${frameworks}/tei/xml/tei/stylesheets/tools/oddbysample.xsl`

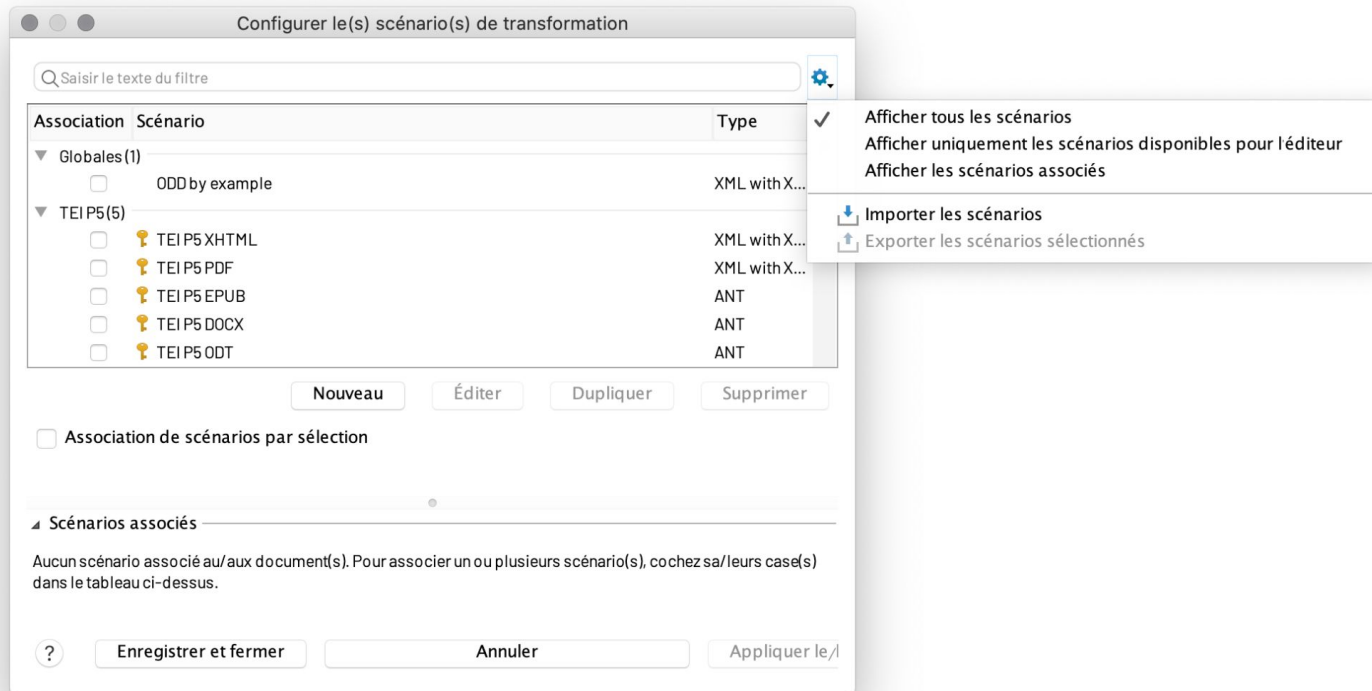
- Sélection le processeur **Saxon 9.xX**



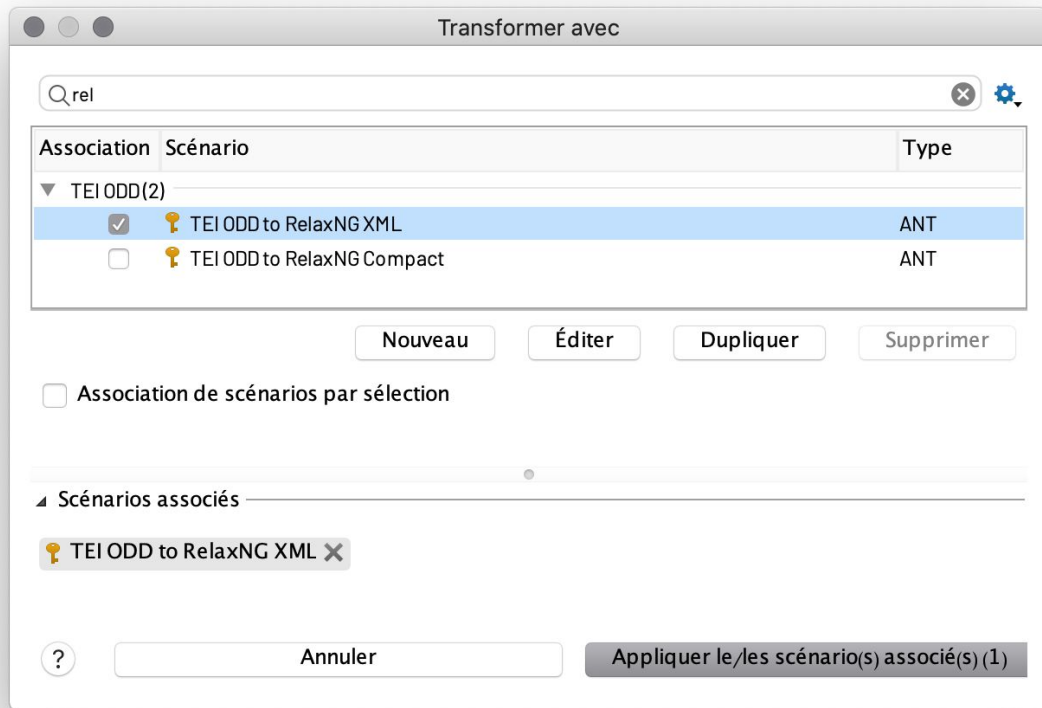
Options avancées > Template (-it) > Remplir le champ avec la valeur “**main**”

- Paramètres > Nouveau > Nom “**corpus**” et Valeur “**\${cfdu}**” (i.e. le répertoire courant)
- Onglet Sortie > Définir un nom de fichier de sortie finissant par .xml pour la future ODD
- Appliquer le scénario au fichier

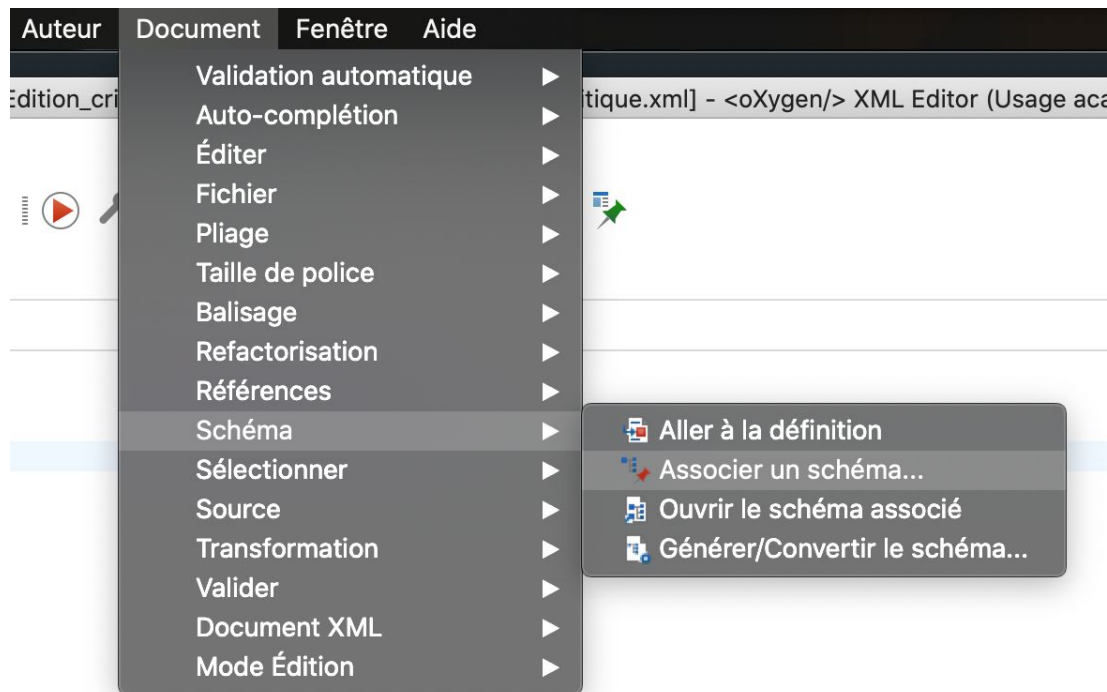
# Afficher tous les scénarios



# Générer un fichier RelaxNG



# Associer un schéma RNG



# Exercice

**Créer une ODD avec  
ODD by example**



# Consigne

Produire une ODD de l'encodage sonnetTEI.xml en suivant le tutoriel ODD by example



# Roma

Roma est un programme permettant de générer automatiquement une ODD grâce à une interface web qui permet de sélectionner les composants de la TEI à utiliser pour l'encodage

[Tutoriel](#)

[Vidéo](#)

**[romabeta.tei-c.org](http://romabeta.tei-c.org)**

# Exercice

**Restreindre l'usage de  
modules avec Roma**

# Consigne

Structurer l'ODD grâce aux modules TEI :

- 1) Créer une ODD pour décrire une pièce de théâtre
- 2) L'appliquer au fichier `Le_Misanthrope_TEI.xml`

# Exercice

A decorative graphic consisting of a light gray square positioned above a blue square, both located on the right side of the slide.

**Restreindre l'usage de  
classes avec Roma**

# Consigne

Structurer l'ODD grâce aux classes TEI :

- 1) Créer une ODD pour une pièce de théâtre
- 2) Modifier la classe `att.global` (en cliquant sur le lien) en supprimant `@n`
- 3) Ouvrir l'ODD, chercher `@n`
- 4) Appliquer le schéma au fichier XML de `Le_Misanthrope_TEI.xml`

# Exercice

A decorative graphic consisting of a light gray square positioned above a blue square, both located on the right side of the slide.

**Restreindre l'usage de  
classes avec Roma**

# Consigne

Structurer l'ODD grâce aux classes TEI :

- 1) Créer une ODD pour une pièce de théâtre
- 2) Modifier la classe att.global en supprimant @n
- 3) Ouvrir l'ODD, chercher @n
- 4) Appliquer le schéma au fichier XML de  
Le\_Misanthrope\_TEI.xml

# Exercice



**Retirer un élément avec  
Roma**



# Consigne

Supprimer un élément :

- 1) Créer une ODD pour une pièce de théâtre
- 2) Supprimer l'élément `speaker`
- 3) Ouvrir l'ODD, chercher `speaker`
- 4) Appliquer le schéma au fichier XML de `Le_Misanthrope_TEI.xml`

# Exercice



**Retirer un attribut avec  
Roma**

# Consigne

Supprimer un élément :

- 1) Créer une ODD pour une pièce de théâtre
- 2) Contraindre les valeurs de l'attribut `type` sur les `div` dans une liste fermée
- 3) Rendre l'attribut obligatoire
- 4) Ouvrir l'ODD, chercher `@type`
- 5) Appliquer le schéma au fichier XML de `Le_Misanthrope_TEI.xml`