

# Humanités numériques : structuration des données et des documents textuels

E. ROUQUETTE

Cours 4 – 12 décembre 2024

## Introduction : de l'encodage au rendu

# Tester son encodage TEI : l'exemple du texte de Flaubert

Le corrigé de l'exercice de la semaine dernière : voir le fichier  
`correction_Flaubert.xml`

# Tester son encodage TEI : l'exemple du texte de Flaubert

Le corrigé de l'exercice de la semaine dernière : voir le fichier `correction_Flaubert.xml`

Il est possible de tester la façon dont un texte encodé pourrait être rendu en HTML sur le site TEI-publisher



<https://teipublisher.com/exist/apps/tei-publisher/index.html>

→ tester avec l'encodage de l'extrait de Flaubert

## Édition scientifique en TEI

# Quelques types d'édition

**Génétique** Axée sur la genèse des textes, leur processus de création.

Exemple : **Madame Bovary** : l'histoire du texte à travers ses brouillons

# Quelques types d'édition

**Génétique** Axée sur la genèse des textes, leur processus de création.

Exemple : **Madame Bovary** : l'histoire du texte à travers ses brouillons

**Diplomatique** Attention à l'aspect visuel du texte et aux spécificités de graphies.

Exemples : **DiScholEd - Éditions scientifiques numériques** : Un ensemble de corpus édités avec version diplomatique et version de lecture des lettres **Montaigne, *Essais*** (1ère édition, 1588)

# Quelques types d'édition

**Génétique** Axée sur la genèse des textes, leur processus de création.

Exemple : **Madame Bovary** : l'histoire du texte à travers ses brouillons

**Diplomatique** Attention à l'aspect visuel du texte et aux spécificités de graphies.

Exemples : **DiScholEd - Éditions scientifiques numériques** : Un ensemble de corpus édités avec version diplomatique et version de lecture des lettres  
**Montaigne, *Essais*** (1ère édition, 1588)

**Fac-similaire** À partir d'images et non de texte

Exemple : **Le *Didascalion* d'Hugues de Saint-Victor**



# Quelques types d'édition

**Génétique** Axée sur la genèse des textes, leur processus de création.

Exemple : *Madame Bovary* : l'histoire du texte à travers ses brouillons

**Diplomatique** Attention à l'aspect visuel du texte et aux spécificités de graphies.

Exemples : DiScholEd - Éditions scientifiques numériques : Un ensemble de corpus édités avec version diplomatique et version de lecture des lettres Montaigne, *Essais* (1ère édition, 1588)

**Fac-similaire** À partir d'images et non de texte

Exemple : *Le Didascalion d'Hugues de Saint-Victor*

**Critique** Mise en parallèle des versions d'un texte pour en révéler les variantes

Exemple : *Cicéron, Timaeus*

## Quelques types d'édition

Pour une réflexion épistémologique sur l'édition numérique, voir par exemple :

Frédéric Duval, « Pour des éditions numériques critiques. L'exemple des textes français », *Médiévales*, 73, automne 2017

# La critique textuelle

- ▶ Une édition critique **compare les manuscrits** pour établir le texte original et rendre compte de cette comparaison dans un **apparat critique**.
- ▶ À partir de la **comparaison des variantes**, on essaie d'établir une **généalogie des textes**, pour donner ensuite un texte le plus proche possible du texte original (archétype).
- ▶ Plusieurs approches sont possibles : approche dite « Lachmanienne », approche dite « Bédiériste ».  
Voir par exemple J.-B. Camps, [Copie, authenticité, originalité dans la philologie et son histoire](#), *Questes*, 29 (2015)

# Quelques étapes pour la constitution d'une édition critique

**Recension** Examen de la tradition d'un texte (manuscrits, *editio princeps*,...)

**Collation** Alignement des différentes versions, pour relever leurs divergences

**Stemma** Arbre généalogique d'un écrit ancien indiquant les relations entre divers manuscrits

Un outil de collation automatique avec python : collatex

[https ://collatex.net/demo/](https://collatex.net/demo/)

# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

► A : Le petit chat est malade

# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche

# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche
- ▶ C : Le petit chien est mort



# La collation pour l'édition critique : exemple fictif

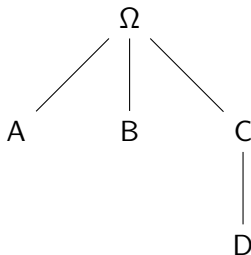
$\Omega$  : Le petit chat est mort

- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche
- ▶ C : Le petit chien est mort
- ▶ D : Le petit chien est décédé

# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

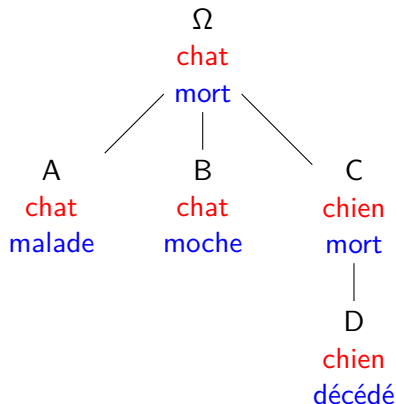
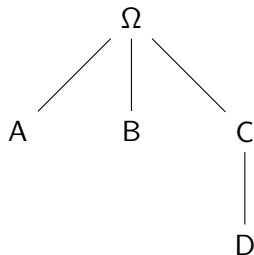
- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche
- ▶ C : Le petit chien est mort
- ▶ D : Le petit chien est décédé



# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

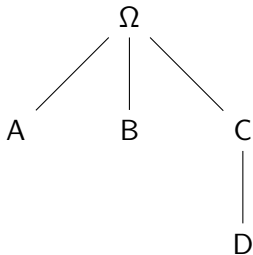
- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche
- ▶ C : Le petit chien est mort
- ▶ D : Le petit chien est décédé



# La collation pour l'édition critique : exemple fictif

$\Omega$  : Le petit chat est mort

- ▶ A : Le petit chat est malade
- ▶ B : Le petit chat est moche
- ▶ C : Le petit chien est mort
- ▶ D : Le petit chien est décédé



O	le petit	chat	est	mort
A	le petit	chat	est	malade
B	le petit	chat	est	moché
C	le petit	chien	est	mort
D	le petit	chien	est	décédé

704 a. Praefatio bibliothecae

2. Praefatio bibliothecae

(Dim. XII, Sic II, 1-2)

1.

Quicquid ab hebreo stiliu Africae atque Latinae  
Sumpit, in hoc totum codice<sup>1</sup>, lector, habes.  
Quo loca prima tenet<sup>2</sup> Genesis, primordia mundi  
Diluviumque canens gestaque magna patrum.  
5 Exodus Aegyptium spoliat, secat aequora rubra.  
Pandit iter heremi : lymphis ibi Lesque datuz.  
Inde sacerdotum gentem Leviticus ornat  
Et typica<sup>3</sup> exponit dona sacris typis.  
Bella utroque liber Numeri describit<sup>4</sup>, et actus  
10 Quis Moyses fregit idola, stupra<sup>5</sup>, duces.  
Post repetit Legem, populo benedicit et alma<sup>6</sup>  
Iura dat, ac dicto carmine victor obit.  
Distulit hinc proprio terram Nauegias heros  
Libro, deletis urbibus atque locis.  
15 Priem iudicibus qui et ordo cucurrit et actus,  
Continuante stilo concinit inde sequeps.  
Mosque sedent quae Ruth, Moabit<sup>7</sup> femina, gressit,  
Clara pia meritis, clara nepote pio.  
Psalmicaneque<sup>8</sup> actus et iniqui gesta Saulis  
20 Prime<sup>9</sup> sequepsque<sup>10</sup> liber, hinc<sup>11</sup>, Samubelis, habes<sup>12</sup>.  
Tertius et<sup>13</sup> quartus, Malachim qui nomen inhestit<sup>14</sup>,  
Regum hebreorum ex ordine gesta canant.  
Hinc<sup>15</sup> sedet<sup>16</sup> Esaias Christum de uirgine nasci  
Qui canit, et gentes ad pia iura uocat.

15a:  $\sigma^{00}F_{1,1}VMBRCln\ d : D_1D_2D_3D_4D_5D_6\ n : 1; N_1N_2N_3N_4N_5N_6N_7\ n : C_1C_2C_3C_4C_5C_6$

17a :  $\sigma^{00}F_{1,1}F_{1,1}VMBRCln\ foci\ exantrati\ [D_1\ n\ pentameter]\ scribitis\ compositi\ Cln\ d$   
Versus Theodoli episcopi de tota Veteris et Novis testamenti instructione G Inquit Theodolius in paudente ff  
Omnes diuine historie ueteris ac noui testamenti libros subsequens metrica praefatio paucis horreolissime  
concludit armetibus [D<sub>1</sub> Versus theodoli episcopi] a Inquit prodigia lucina ueteris historie c Versus in  
fontes biblicum quae descripti sunt. Sic. Unde  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : compositi G Inquit  
 $\sigma^{00}VMBRCln\ d\ n\ c\ Sic$  : uerit  $F_{1,1}$  "primordia  $\sigma^{00}VMBRCln\ d\ n\ c\ Sic$  : uerit  $F_{1,1}$  "typica [14p]  
 $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : quae  $\sigma^{00}F_{1,1}$  "describit  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : descripti At designat V "aliqua  
 $\sigma^{00}VMBRCln\ d\ n\ c\ Sic$  : aliquo Cln uerit  $F_{1,1}$  "alma  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : uerit VRO "suadita  
 $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : Alitudo G "psalmicaneque  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : psalmicaneque G  
"prime  $\sigma^{00}F_{1,1}F_{1,1}Cln\ n\ c\ Sic$  : prima RG d non legitur  $F_{1,1}$  "sequepsque  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : se  
quens c "hinc  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : hic VRO "habes  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : habes n "et  
 $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : ac n "nomine inhestit  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : nomine gressit c  
"hinc  $\sigma^{00}VMBRCln\ d\ n\ c\ Sic$  : hic  $F_{1,1}$  "sedet  $\sigma^{00}F_{1,1}VMBRCln\ d\ n\ c\ Sic$  : et G

Dim. XII, Sic II, 1-2 705

2. Préface à la Bible

(Dim. XII, Sic II, 1-2)

1.

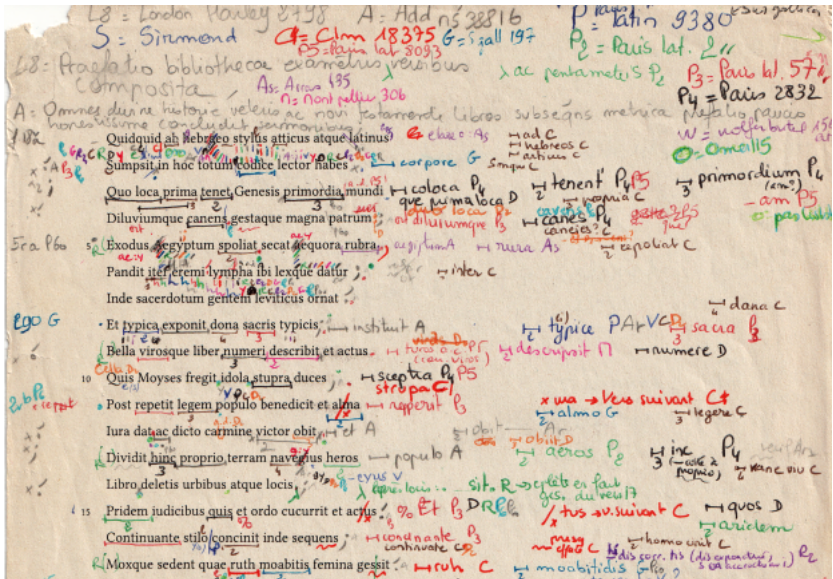
Tout ce qu'à l'hébreu les stylets grec et latin  
Ont emprunté, tu l'as en entier dans ce livre, lecteur.  
La Genèse y occupe la première place, chantant les premiers moments du monde<sup>1</sup>,  
Le Déluge<sup>2</sup> et les hauts faits des Pères<sup>3</sup>.  
L'Exode dépouille l'Egypte<sup>4</sup>, divise la mer Rouge<sup>5</sup>,  
5 Ouvre le chemin du désert<sup>6</sup> : là sont données l'eau et la Loi<sup>7</sup>.  
Le Lévitique ensuite pure la lignée des prêtres<sup>8</sup>.  
Et expose les offrandes typologiques pour les sacrifices typologiques<sup>9</sup>.  
Le livre des Nombres décrit les guerres et les hommes, et les actes  
Par lesquels Moïse brisa les idoles, les impures, les chefs.  
10 Après quoi, il va chercher la Loï, bérut le peuple, et lui donne  
Les doux commandements<sup>10</sup>, puis meurt, victorieux, après avoir dit un cantique<sup>11</sup>.  
Le héros fils de Nün<sup>12</sup> partage ensuite dans son livre la terre<sup>13</sup>.  
Après la destruction des villes et des campagnes<sup>14</sup>.  
15 Le livre suivant chante ensuite, sans que s'arrête son stylet,  
L'ordre et les actes des juges d'autrefois.  
Se trouve ensuite ce que fit Ruth, la Moabite,  
Illustre pour ses pieux mérites, illustre pour sa pieuse descendance<sup>15</sup>.  
Puis ce sont les actes du Psalmit<sup>16</sup> et les faits de l'unique Saül  
Que vous contenez, premier et second livres de Samuel.  
20 Ses troisième et quatrième livres, auxquels est attaché le nom de Malachim<sup>17</sup>,  
Chantent dans l'ordre les faits des rois des hébreux.  
Puis se trouve Isaïe, qui chante que le Christ naît de la Vierge<sup>18</sup>  
Et qui appelle les nations à suivre les pieux commandements.

<sup>1</sup>Gen 1-11 <sup>2</sup>Gen 6-8 <sup>3</sup>Gen 12-50 <sup>4</sup>Ex 3, 12 <sup>5</sup>Ex 14, 15-21 <sup>6</sup>Ex 15, 22 <sup>7</sup>Ex 17, 1-7 <sup>8</sup>Ex 18-24 <sup>9</sup>Ex 25-30  
<sup>10</sup>Ex 17, 12-13 <sup>11</sup>Ex 15-18 <sup>12</sup>Ex 31, 30-34 <sup>13</sup>Num 13-21 <sup>14</sup>Num 31 <sup>15</sup>Ex 15-18 <sup>16</sup>Ps 118 <sup>17</sup>2 Sam 1-12

<sup>18</sup>Isaïe. "Ruth donna naissance à Chel, père de Boaz, père de David, de qui descend Jésus. "David, à qui  
sont attribués les Psalmes. "Mal qui signifie « des rois » en hébreu, selon Isidore de Séville, Etymologiae,  
VI, 11, 11.

21 Jérôme, Epistolas, 53, 8 : « Malachias, id est tertius et quartus Regum liber » 22 Isidore de Séville, Etymologiae, VI, n. 11 : « Malachias liber primus appellatur, eo quod mores Iudae et israeliticus gentis gentisque  
civium per illum agerit temperant » 23 Jérôme, Epistolas, 53, 8 : non prophetam nisi uideret Ierosolam, sed euangelium »

## Une collation manuelle...



# Encodage XML

```
<l>Quo loca prima <app>
  <rdg type="source"
    wit="#AM #reglat2078 #montpellier306 #angel1515,
    #gall197 #clm18375 #D #N #Co #Si">
    tenet</rdg>
    <rdg wit="#8093 #2832">-nent</rdg>
  </app> Genesis, <app>
    <rdg type="source" wit="AM #reglat2078 #montpellier306,
    #angel1515 #gall197 #clm18375 #D #N #Co #Si">
    primordia</rdg>
    <rdg wit="#8093 #2832">-iam</rdg>
  </app> mundi
</l>
```

# Transformation vers L<sup>A</sup>T<sub>E</sub>X du texte encodé en XML : XSLT

Un bout de code XSLT :

```
<!-- appeler les variantes et le texte choisi -->
<xsl:template
  ↪  match="tei:TEI/tei:text/tei:body/tei:l/tei:app">
  \var{<xsl:call-template name="source"/>
  <xsl:call-template name="sourceWit"/>}
  {<xsl:call-template name="variantes"/>}
  {<xsl:call-template name="omissions"/>}
</xsl:template>
```



# Résultat pdf : l'apparat

---

*Tit. : sine tit.  $\Theta^{AM}F_1F_2$  Praefatio bibliothecae (-ce) exametris [D<sub>1</sub> ac pentametris] uersibus composita Clm d*  
*Versus Thietolfi episcopi de tota Veteris et Noui testamenti instructione G Incipit Theodulfus in pandecten R*  
*Omnes diuinae historiae ueteris ac noui testamenti libros subsequens metrica prefatio paucis honestissime*  
*concludit sermonibus [N<sub>4</sub> Versus theudulfi episcopi] n Incipit prologus tocius ueteris historie c Versus*  
*in fronte Bibliorum quae ipse describi fecit Sir. <sup>1</sup>codice  $\Theta^{AM}F_1F_2VMtRClm d n c$  Sir. : corpore G <sup>2</sup>tenet*  
 *$\Theta^{AM}VMtRGClm d n c$  Sir. : -nent  $F_1F_2$  <sup>3</sup>primordia  $\Theta^{AM}VMtRGClm d n c$  Sir. : -iam  $F_1F_2$  <sup>4</sup>typica (tip-)*  
 *$\Theta^AF_1F_2VMtRGClm d n$  Sir. : -pice  $\Theta^M c$  <sup>5</sup>describit  $\Theta^{AM}F_1F_2RGClm d n c$  Sir. : descripsit Mt designat V <sup>6</sup>stupra*  
 *$\Theta^{AM}VMtRG d n c$  Sir. : strupa Clm scepra  $F_1F_2$  <sup>7</sup>alma  $\Theta^{AM}F_1F_2MtClm d n c$  Sir. : -mo VRG <sup>8</sup>moabit*  
 *$\Theta^{AM}F_1F_2VMtRClm d n c$  Sir. : -bitidis G <sup>9</sup>psalmicanique  $\Theta^{AM}F_1F_2VMtRClm d n c$  Sir. : psalmigrafique G*  
*<sup>10</sup>prime  $\Theta^{AM}F_1F_2Clm n c$  Sir. : prima RG d non legitur  $F_1$  <sup>11</sup>sequensque  $\Theta^{AM}F_1F_2VMtRGClm d n$  Sir. : se-*  
*quens c <sup>12</sup>hinc  $\Theta^{AM}F_1F_2MtRGClm d n$  Sir. : hic Vfo c <sup>13</sup>habes  $\Theta^{AM}F_1F_2VMtRGClm d c$  Sir. : habe n <sup>14</sup>et*  
 *$\Theta^{AM}F_1F_2VMtRGClm d c$  Sir. : ac n <sup>15</sup>nomen inhesit (-haesit)  $\Theta^{AM}F_1F_2VMtRGClm d n$  Sir. : nomine gessit c*  
*<sup>16</sup>hinc  $\Theta^{AM}VMtRGClm d n c$  Sir. : hic  $F_1F_2$  <sup>17</sup>sedet  $\Theta^{AM}F_1F_2VMtRClm d n c$  Sir. : et G*

# Le module textcrit pour les édition critiques

Module pour produire des apparats critiques : [Critical Apparatus](#)  
→ description d'éléments qui permettent de produire des apparats critiques pour des éditions scientifiques

# La méthode d'encodage choisie (→ métadonnées)

## ` Les deux méthodes d'encodage des variantes

Méthode d'encodage à indiquer dans `<encodingDesc>` :

**1. double-end point** : le lemme est indiqué dans le texte directement et les variantes à part

```
<variantEncoding method="double-end-point"  
location="internal"/>
```

**2. Parallel segmentation** : toutes les variantes sont signalées au fil du texte.

```
<variantEncoding method="parallel-segmentation"  
location="internal"/>
```

→ celle que nous allons utiliser. Elle permet d'obtenir un appareil positif, ou encore d'extraire le texte complet de n'importe quel témoin.

# La recension des témoins (→ métadonnées)

- Déclaration des témoins dans l'élément `<sourceDesc>` avec les éléments `<listWit>` et `<witness>`

```
<sourceDesc>
  <listWit>
    <witness xml:id="A">
      Paris, BnF, Latin 8093
    </witness>
    <witness xml:id="B">
      Paris, BnF, Latin 2832
    </witness>
  </listWit>
</sourceDesc>
```

# Exercice : Encoder un court texte avec appareil critique 1/ Recenser les témoins

Nous allons encoder en vue d'une édition critique le *Carmen XX* de Sidoine Apollinaire.

1. Indiquer la méthode d'encodage dans `<encodingDesc>` :  
`<variantEncoding method="parallel-segmentation" location="internal"/>`
2. Entrer dans le `<sourceDesc>`, au sein de `<listWit>`, la liste des manuscrits fournies dans le fichier `transcription_Sidoine.txt`.  
Chaque manuscrit correspond à une entité `<witness>` et a un identifiant (`@xml :id=`)

**nb :** Nous nous appuyons pour cet exercice sur l'édition d'André Loyer (Les Belles Lettres, 1961).

# Encoder une variante simple

## Encoder les variantes :

- `<app>` Une entrée dans l'apparat critique (→un lieu variant). Doit contenir au moins une leçon
- `<lem>` Le lemme (la leçon choisie)
- `<rdg>` Une leçon, à l'intérieur d'une entrée dans l'apparat
- `<rdgGrp>` Regroupe deux leçons ou plus qui sont perçues comme ayant une relation (exemple : des variantes orthographiques d'un même mot)

## Indiquer les témoins :

`<rdg wit="#Exemple">...</rdg>`

# Encoder une variante simple

Exemple :

Le petit <app>

<lem wit="#A #B">chat</lem>

<rdg wit="#C #D">chien</rdg>

</app> est mort

# Exercice : Encoder un court texte avec appareil critique

## 2/Encoder les variantes

- ▶ Encoder le texte de Sidoine avec `<head>` et `<l>`
- ▶ Encoder les variantes du texte avec les éléments `<app>`, `<lem>`, `<rdg>` et l'attribut `@wit`

→ Lorsque vous tapez l'attribut `@wit`, oXygen vous propose comme valeur chaque témoin présent dans `<listWit>`, en indiquant quel témoin a déjà été noté

- ▶ Tester votre solution sur le site **TEI Critical Apparatus Toolbox** (<http://teicat.huma-num.fr/>)





# Ajouter des informations

Quelques attributs possibles au sein des éléments `<lem>` ou `<rdg>` :

`@type` Type de variante. Exemple de valeur : `"substantive"`, `"semantic"`, `"orthographic"`, `"omission"`

`@hand` main responsable de la variante. Renvoie à un `@xml:id` défini dans un `<handNote>` lui-même contenu dans `<handDesc>`.

Exemple : `<handNote xml:id="scribe1">Premier scribe</handNote>`

`@resp` référence à l'`xml:id` du responsable (l'auteur de la correction,)

`@cert` le degré de certitude (`"high"`, `"low"`)

etc...

# Informations sur les témoins

`<witDetail>` Détails sur les témoins

`<witStart>`, `<witEnd>` Témoin fragmentaire

`<lacunaStart/>`, `<lacunaEnd/>` Début et fin de lacune d'un  
manuscrit

Exemple :

Le petit `<app>`

```
<lem wit="#A #B">chat</lem>
```

```
<rdg wit="#C #D">chien</rdg>
```

```
<rdg wit="E" type="omission"/>
```

```
<witDetail wit="#E">Le mot a été remplacé par un
```

```
  ↪ dessin</witDetail>
```

```
</app> est mort
```

## Exercice : Encoder un court texte avec appareil critique

### 3/Encoder les variantes (fin)

Encoder les variantes du titre, en indiquant l'omission au moyen de l'attribut @type dans une balise <rdg/> auto-fermante

## Aller plus loin

<https://teibyexample.org/exist/tutorials/TBED07v00.htm#apparatus>

<http://developpements.enc.sorbonne.fr/diple/schema/witness/>

# Autres outils pour des éditions scientifiques :

## Le module transcr pour les facs-similés

Module pour représenter les sources primaires : [Representation of Primary Sources](#)

- ▶ indications sur la matérialité des documents
- ▶ permet d'encoder les images sources encodées pour un alignement texte/image → combiner fac-similé et transcription

Exemple : manuscrit interactif sur le site de l'EnC :

<http://theleme.enc.sorbonne.fr/dossiers/vue97.php>

Outil pour annoter une image en TEI :

<http://teicat.huma-num.fr/zoner/index.php#catList>

# Autres outils pour des éditions scientifiques :

## Au sein du module core

`<lb/>`, `<cb/>`, `<pb/>` Indiquer un début de ligne, de page ou de colonne

`<choice>` Traiter une même donnée de plusieurs manières :

`<orig/>`/`<reg>` Version originale/version régularisée

`<sic/>`/`<corr>` Version fautive/version corrigée

`<abbr/>`/`<expn>` Abréviation/résolution de l'abréviation

`<add>` (ajout) contient un ajout

`<del>` (suppression) contient une séquence supprimée ou indiquée comme telle

`<gap>` (omission) omission pour raison éditoriale ou pour toute autre raison

`<hi>` (mise en évidence) désigne une séquence mise en évidence d'un point de vue graphique

`<unclear>` (incertain) contient une séquence difficile à transcrire

# Autres outils pour des éditions scientifiques :

## Encoder des caractères spécifiques

### Les entités

Commencent par le signe `&` et se terminent par `;` Permettent d'encoder des objets spécifiques, sous forme d'alias :

- ▶ des caractères spéciaux (caractères réservés en XML) :

`&lt;` `<`

`&amp;` `&`

- ▶ des caractères que l'on déclare dans une DTD et qu'il sera possible d'utiliser dans le document XML

Liste des entités : [https://en.wikipedia.org/wiki/List\\_of\\_XML\\_and\\_HTML\\_character\\_entity\\_references](https://en.wikipedia.org/wiki/List_of_XML_and_HTML_character_entity_references)

# Autres outils pour des éditions scientifiques :

## Encoder des caractères spécifiques

### Exemple d'utilisation d'entité

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI [
  <!ENTITY s-long "f">
]>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>...</teiHeader>
  <text>
    <body>
      <p>Le &s-long;ene&s-long;chal du roy</p>
    </body>
  </text>
</TEI>
```

Voir le fichier exemple-tei.xml (à tester sur TEI publisher)



# Autres outils pour des éditions scientifiques :

## Encoder des caractères spécifiques

### Outils pour encoder les caractères médiévaux

MUFI [Medieval Unicode Font Initiative](#) → documentation sur l'encodage unicode des caractères médiévaux

module Gaiji : [Characters, Glyphs, and Writing Modes](#) → permet l'utilisation des caractères autres que les standards de l'alphabet latin

Personnaliser la TEI : schémas et ODD

# Les modèles TEI

Aucun projet n'utilise toutes les balises possibles. Il existe des modèles (schémas) sélectionnant un certain nombre d'éléments.

Exemple :

`tei_lite` Couvre les besoins de 90% des utilisateurs

`tei_corpus` pour les corpus linguistiques

`tei_ms` pour la description de manuscrits

`tei_drama` pour les corpus dramatiques

Toutefois, ces schémas ne conviennent pas toujours. → Il est possible de **générer son propre schéma TEI** avec **Roma** :

<https://roma.tei-c.org/>. Permet aussi documenter son travail, de préciser ses choix.

## One Document Does it all

- ▶ Un ODD est un fichier XML
- ▶ Ce fichier permet de produire à la fois un **schéma**, lisible par la machine, et une **documentation**, lisible par l'humain, des balises et attributs contenu dans le schéma et de leurs relations
- ▶ → **validation** du document, **guidage** de l'encodage

## One Document Does it all

Quelques éléments d'un ODD :

`<schemaSpec>` élément racine

`<elementSpec>` définit un élément avec tous ses attributs possibles  
ou avec une partie seulement (personnalisation)

`<attList` liste des attributs associés à l'élément

`<elementRef>` appelle la définition existante d'un élément

`<classRef>` Fait appel à une classe existante

`<moduleRef>` Fait appel à un module existant

...

## One Document Does it all

Exemple d'élément personnalisé :

```
<elementSpec ident="app" mode="change">  
  <attList>  
    <attDef ident="from" mode="delete">  
      <attDef ident="to" mode="delete">  
        <attRef name="rend" class="att.global.rendition"/>  
      </attList>  
    </attList>  
  </elementSpec>
```

# Produire un ODD avec ROMA

<https://roma.tei-c.org/>

## Choix de départ

- ▶ Modèle TEI existant. Exemples :
  - ▶ TEI-all (on enlèvera des éléments)
  - ▶ TEI minimal (on ajoutera des éléments)
- ▶ ODD vierge
- ▶ ODD déjà existant

À partir de ce point de départ, il est possible d'ajouter des éléments un à un, ou de sélectionner tous les éléments d'un module

# Exercice : Produire son ODD avec ROMA (1)

1. Créer avec oXygen un document utilisant comme schéma le modèle TEI-drama :  
Nouveau - modèle du framework - TEI P5 - Drama
2. Y coller l'encodage des rôles du Cid (cours n.3, fichier `solution_Cid.xml`)
3. Repérer le problème : l'attribut `@rend` n'est pas accepté dans l'élément `<castGroup rend="braced">`

→ Nous allons produire étape par étape un ODD ajoutant simplement au module drama l'attribut `@rend` pour l'élément `<castGroup>`



## Exercice : Produire son ODD avec ROMA (1)

**Créer un ODD avec Roma à partir de TEI-drama** (lui donner un nom et un identifiant). Puis :

1. cliquer sur l'élément `<castGroup>` → essayer de lui ajouter l'attribut `@rend` : celui-ci ne fait pas partie de la customization.
2. Chercher la classe d'attribut dans laquelle se trouve `@rend` : `global.rendition`. Rajouter la classe.
3. Au sein de cette classe, ne sélectionner que l'attribut `@rend`
4. Dans l'élément `<castGroup>`, rajouter `@rend` et l'éditer : choisir « Default (Optional) »



# Associer l'ODD au document XML

1. à la fin de l'édition de l'ODD, télécharger le résultat :  
Download- Customization as ODD
2. mettre l'ODD dans le dossier où se trouve le document XML
3. ouvrir l'ODD dans oXygen
4. cliquer sur la clé à molette pour « Configurer le scénario de transformation »



# Associer l'ODD au document XML





Une fenêtre s'ouvre :

Association	Scénario	Type
TEI ODD (9)		
<input type="checkbox"/>	TEI ODD XHTML	XML with XSLT
<input checked="" type="checkbox"/>	TEI ODD PDF	XML with XSLT
<input type="checkbox"/>	TEI ODD EPUB	ANT
<input type="checkbox"/>	TEI ODD DOCX	ANT
<input type="checkbox"/>	TEI ODD ODT	ANT
<input checked="" type="checkbox"/>	TEI ODD to RelaxNG XML	ANT
<input type="checkbox"/>	TEI ODD to DTD	ANT
<input type="checkbox"/>	TEI ODD to XML Schema	ANT
<input type="checkbox"/>	TEI ODD to RelaxNG Compact	ANT
Projet (2)		
<input type="checkbox"/>	Flowers sample to WebHelp Responsive (Flowers template)	DITA-OT
<input type="checkbox"/>	Garage sample to WebHelp Responsive (Mechano template)	DITA-OT

☐ Association de scénarios par sélection

4 Scénarios associés

 TEI ODD to RelaxNG XML   TEI ODD PDF 

## Associer l'ODD au document XML

- ▶ Sélectionner « TEI ODD to RelaxNG XML »
- ▶ (et « TEI ODD to PDF » pour produire la documentation)
- ▶ cliquer sur « Appliquer le/les scenarios associé(s) »
- ▶ On obtient un dossier `out` dans lequel se trouve le schéma généré ; l'associer à son fichier XML au moyen de l'épingle rouge



→ il est désormais possible d'ajouter l'attribut `@rend`

# Exercice pour la semaine prochaine : : Produire son ODD avec ROMA (2) - partir du modèle minimal

## Créer un ODD pour générer un schéma convenant à notre encodage du Dictionnaire des idées reçues de Flaubert

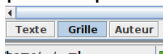
- ▶ Dans Roma, partir du modèle « TEI minimal ». Donnez un titre et un identifiant à votre personnalisation
- ▶ Ajouter les différents éléments utilisés pour l'encodage du Dictionnaire
- ▶ Une fois votre schéma obtenu, vérifiez-le (si vous l'associez au document XML, celui-ci doit-être valide)
- ▶ **Suite de l'exercice** : encoder la fin de la première partie du dictionnaire avec ce schéma – Voyez si vous devez compléter votre schéma

# Exercice pour la semaine prochaine : : Produire son ODD avec ROMA (2) - partir du modèle minimal

## Créer un ODD pour générer un schéma convenant à notre encodage du Dictionnaire des idées reçues de Flaubert

### Conseils pour cet exercice

- ▶ Partir du mode « grille » dans l'édition du document XML pour repérer tous les éléments utilisés



- ▶ Testez d'abord votre schéma en ayant ajouté seulement des éléments – pour repérer s'il faut ou non ajouter des attributs
- ▶ Pensez à chercher par modules pour les éléments propres au dictionnaire
- ▶ Faites des exports intermédiaires sous forme d'ODD pour éviter de perdre votre travail