TP2 : oXygen

TEI@Oxford *Sept 2012*

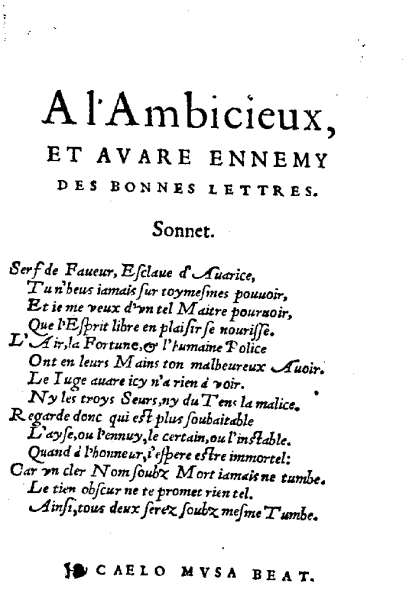
1 À la découverte d'Oxygen - 2

Dans ce deuxième exercice, nous allons approfondir votre expérience avec Oxygen pour :

- l'encodage XML d'un document existant
- la représentation des structures aperçues dans un texte littéraire
- l'affichage en mode « auteur »

2 Sample text

Dans cet exercice, nous allons baliser ce sonnet de Du Bellay, un écrivain du XVIe siècle :



A l'Ambicieux,
ET AVARE ENNEMY
DES BONNES LETTRES.

Sonnet.

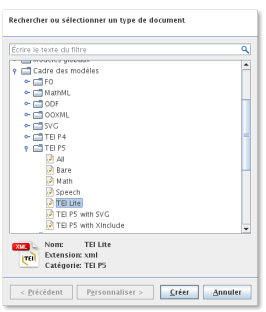
Serf de Faueur, Eſclaue d'Auarice,
Tu n'beus iamais ſur toymeſmes pouuoir,
Et ie me veux d'vn tel Maitre pouruoir,
Que l'Eſprit libre en plaiſir ſe nouriſſe.
L'Air, la Fortune, & l'humaine Police
Ont en leurs Mains ton malheureux Auoir.
Le Iuge auare icy n'a rien à voir.
Ny les troys Seurs, ny du Tens la malice.
Regarde donc qui eſt plus ſoubaitable
L'ayſe, ou l'ennuy, le certain, ou l'inſtable.
Quand à l'honneur, i'eſpere eſtre immortel:
Car vn cler Nom ſoubz Mort iamais ne tumbe.
Le tien obſcur ne te promet rien tel.
Ainſi, tous deux ſerez ſoubz meſme Tumbe.

CAELO MVSA BEAT.

Pour vous simplifier la vie, on a déjà préparé une transcription prête à l'emploi... ne la regardez pas tout de suite !

3 Création du document

Ouvrez Oxygen. Cliquez sur l'icône Nouveau en haut à gauche, (ou sélectionnez Nouveau du menu Fichier, ou tapez CTRL-N.) pour ouvrir la boîte de dialogue Nouveau

- Cette fois, choisissez Cadre des modèles, ensuite TEI P5, et ensuite TEI-Lite, pour sélectionnez un schéma TEI plus complet.
- 

- Comme auparavant, Oxygen vous propose un squelette du document ; vous observerez que le schéma TEI lite décrit un document TEI plus élaboré avec un choix de balises plus important.
- Complétez l'entête TEI avec des indications de titre ('Un sonnet du Du Bellay numérisé'), distribution, et source ('encodé à partir d'un facsimilé de la première édition produit au CESR, Tours')
- Enlevez toute la partie proposée par défaut pour le `<body>` (mais retenez les balises `<body>` et `</body>`).

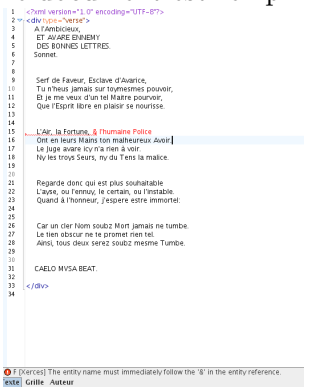
4 Ajout de texte au document

Rien ne vous empêche de taper à la main tout le poème directement. Mais pour gagner du temps on vous propose la démarche suivante :

- Assurez-vous que le curseur soit toujours entre les deux balises `<body>` et `</body>` de votre document vide
- Dans le menu Document, sélectionnez Fichier, et ensuite Insèrer un fichier
- Naviguez jusqu'au fichier duBellay.txt dans le dossier Travaux qui devrait se trouver sur votre Bureau, et insérez-le. Ce fichier est aussi disponible sur le web, à cet URL)

5 Attention, Will Robinson !

Votre document est rempli de taches rouges!



On va régler cela petit à petit. Commençons avec la partie la plus sérieuse : où le texte lui-même est affiché en rouge.

- Oxygen vous indique que la partie en rouge n'est pas bien formée
- Il y a un message en bas de l'écran qui vous aidera (un peu) à identifier le problème. Sauriez vous comment le résoudre ?
- Conseil : souvenez-vous du fait que l'esperluette est un caractère magique en XML, qui doit donc être représentée de manière indirecte.

6 Structuration du document

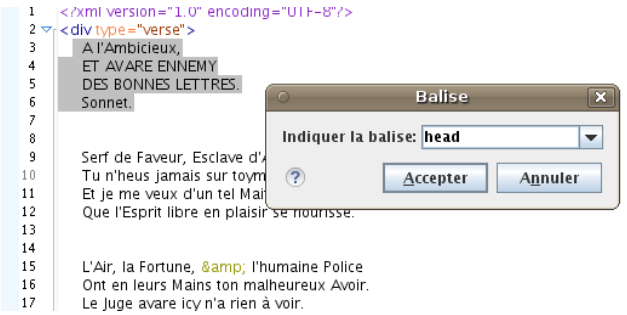
Il est possible d'identifier dans ce document plusieurs sous-parties. En particulier, il contient :

- un titre ... à baliser avec la balise `<head>`
- quatre strophes ... à baliser avec la balise `<lg>`
- des vers ... à baliser avec la balise `<l>`

- une citation d'Horace ... à baliser avec la balise <quotte>

Allons-y !

- Avec la souris, sélectionnez les quatre premières lignes du document 'À l'ambitieux et avare ... sonnet '; ceux-ci constituent son titre.
- Tapez CTRL-E (ou sélectionnez XML Refactoring et ensuite Entourer des balises dans le menu Document
- Oxygen vous propose toutes les balises disponibles à cet emplacement.
- Sélectionnez <head> et cliquez Accepter



- Ce sonnet est bicéphale : coupez en deux parties le <head> existant.. La première partie est le titre. La seconde partie est le mot sonnet. Chacune de ces deux parties devrait-être balisée avec un <head> au début et à la fin, en vous plaçant avant ou après le mot 'Sonnet' et en vous servant de CTRL-ALT-D ou d'une autre manière comme lors du premier exercice.

7 Balisage du poème

Enfin, il faut travailler un peu sur l'étiquetage du poème lui-même. Notre but serait de baliser chaque vers, se servant de la balise TEI <l> ('line') et en amont de baliser chaque ensemble de vers constituant une strophe avec la balise TEI <lg>.

Bien sur, il y a plusieurs manières de faire. On va vous décrire une méthode assez rapide, mais si vous préfériez une autre démarche, on n'insiste pas !

- Sélectionnez avec la souris tout le texte du poème, jusqu'à la citation d'Horace 'caelo musa beat'.
- Entourez toute le poème d'une balise <l>.
- Entourez la citation d'Horace d'une balise <quote>
- Le petit carré vert apparaît ! Avons-nous terminé ? Hélas non : un document peut être valide, tout en contenant des mensonges ! Voyez-vous des mensonges ? Cliquez sur le bouton Indentation (ou tapez CTRL-SHIFT-P).

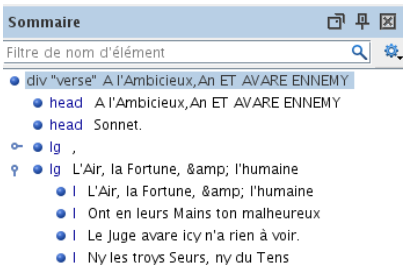
```
<body>

  <head>A l'Ambicieux, ET AVARE ENNEMY DES BONNES LETTRES. Sonnet.</head>

  <l>Serf de Faveur, Esclave d'Avarice, Tu n'heus jamais sur toymesmes pouvoir, Et je me veux
    d'un tel Maitre pourvoir, Que l'Esprit libre en plaisir se nourisse. L'Air, la Fortune,
    &amp; l'humaine Police Ont en leurs Mains ton malheureux Avoir. Le Juge avare icy n'a rien à
    voir. Ny les troys Seurs, ny du Tens la malice. Regarde donc qui est plus souhaitable
    L'ayse, ou l'ennuy, le certain, ou l'instable. Quand à l'honneur, j'espere estre immortel:
    Car un cler Nom soubz Mort jamais ne tumbe. Le tien obscur ne te promet rien tel. Ainsi,
    tous deux serez soubz mesme Tumbe. </l>

  <quote> CAELO MVSA BEAT.</quote>

</body>
```

Parce qu'elles ne sont pas explicités par le balisage, la distinction entre chaque vers et entre chaque strophe ne sont plus affichées. Il fait donc les baliser. (Si vous ne voyez pas bien où indiquer ces séparations, vous pouvez ré-établir l'affichage originel en tapant CTRL-Z)

- Mettez le curseur au début de chaque vers, et distinguez-le du précédent, en vous servant du bouton division ou en tapant ALT-SHIFT-D
- Si cela vous semble plus rapide, vous pouvez aussi 'tricher' : copiez la séquence `</l><l>`, et collez-la au commencement de chaque vers.
- Quand vous aurez fini, cliquez encore sur le bouton Indentation. Ou sont les strophes ?
- Entourez tous les vers (tous les `<l>`) avec une balise `<lg>`.
- Divisez cette strophe unique en quatre strophes distinctes. Attention : il faut effectuer cette distinction au bon endroit — entre la fin d'un vers et le début du suivant, (i.e. après une balise fermante et avant une balise ouvrante).

8 Affichage du texte balisé

Est-ce que la disparition graduelle de notre beau sonnet au dessus d'un nuage de balises commence à vous inquiéter un peu ? Cela nous permet au moins de visualiser la structure émergeante du sonnet – regardez par exemple dans la fenêtre 'Sommaire' à gauche :



Pour contrôler l'affichage d'un text balisé, il faut préciser comment nous désirons visualiser chaque balise. Ces précisions se font avec ce qu'on appelle une *feuille de style*.

- Sélectionnez XML Document et ensuite Associer une feuille de style à partir du menu Document.
- Naviguez jusqu'au fichier duBellay.css dans le dossier Travaux qui devrait se trouver sur votre Bureau, et associez-le (ce fichier est aussi disponible à l'URL sur le web).
- Notez qu'une ligne contenant un "processing instruction" est ajouté au début de votre document.
- Contrôlez l'effet sur l'affichage de votre document au mode Auteur.

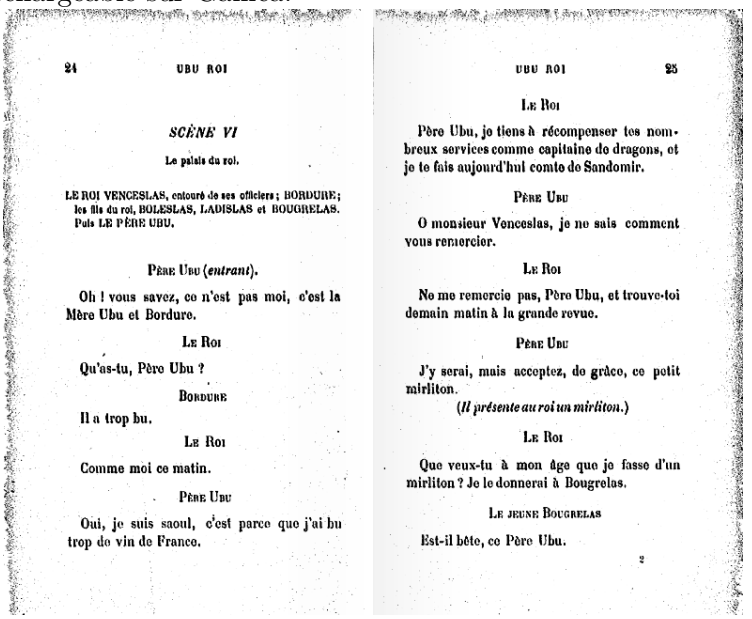
On peut changer cet affichage. Ouvrez le fichier duBellay.css avec Oxygen et modifiez-le, si cela vous intéresse. Mais il y a des limites :

- CSS n'est pas un langage XML, et ne permet que de modifier l'affichage du contenu d'un document XML.
- On peut changer la taille de police, l'étendu des marges, les couleurs etc.
- Changez la couleur du titre de `red` en `purple.`

Des transformations beaucoup plus intéressantes sont possibles avec un autre langage de feuille de style : le XSLT.

9 Par ma chandelle verte...

Voici le commencement d'une scène extraite de *Ubu Roi* de Alfred Jarry (1896), dans la version téléchargeable sur Gallica.



24 UBU ROI

SCÈNE VI

Le palais du roi.

LE ROI VENCESLAS, entouré de ses officiers ; BORDURE ; les fils du roi, BOLESLAS, LADISLAS et BOUGRELAS. Puis LE PÈRE UBU.

PÈRE UBU (*entrant*).

Oh ! vous savez, ce n'est pas moi, c'est la Mère Ubu et Bordure.

LE ROI

Qu'as-tu, Père Ubu ?

BORDURE

Il a trop bu.

LE ROI

Comme moi ce matin.

PÈRE UBU

Oui, je suis saoul, c'est parce que j'ai bu trop de vin de France.

UBU ROI 25

LE ROI

Père Ubu, je tiens à récompenser tes nombreux services comme capitaine de dragons, et je te fais aujourd'hui comte de Sandomir.

PÈRE UBU

O monsieur Venceslas, je ne sais comment vous remercier.

LE ROI

Ne me remercie pas, Père Ubu, et trouve-toi demain matin à la grande revue.

PÈRE UBU

J'y serai, mais acceptez, de grâce, ce petit mirliton.

(*Il présente au roi un mirliton.*)

LE ROI

Que veux-tu à mon âge que je fasse d'un mirliton ? Je le donnerai à Bougrelas.

LE JEUNE BOUGRELAS

Est-il bête, ce Père Ubu.

Et voici le début de la même scène, transcrite avec Word:

Scène VI

Le palais du roi.

LE ROI VENCESLAS, entouré de ses officiers; BORDURE; les fils du roi, BOLESLAS, LADISLAS & BOUGRELAS. Puis UBU.

Père Ubu (*entrant*):

Oh! vous savez, ce n'est pas moi, c'est la mère Ubu et Bordure.

Le Roi:

Qu'as-tu, Père Ubu?

Bordure:

Il a trop bu.

Le Roi:

Comme moi ce matin.

Père Ubu:

Oui, je suis saoul, c'est parce que j'ai bu trop de vin de France.

Le Roi:

Père Ubu, je tiens à récompenser tes nombreux services comme capitaine de dragons, et je te fais aujourd'hui comte de Sandomir.

10 Transformation d'un fichier Word

Vous avez vu comment baliser un fichier « text ». Comment pouvons-nous faire de même avec notre fichier Word ?

Nous pourrions exporter le fichier Word au format « plein text ». Nous pourrions aussi faire un copier-coller. Mais dans chaque cas, nous perdrions le formatage qui distingue (par exemple) le nom de chaque locuteur de ces énoncés.

Si le fichier Word est bien stylé, on peut le transformer en XML, sans perte d'information. L'outil OxGarage nous aidera.

- Ouvrez votre Internet browser, et allez sur le site http://www.tei-c.org/ege-webclient/
- Cliquez sur Text Documents. Dans la liste des formats de saisi proposée, cochez la case pour Microsoft Word Document.

- Une liste des formats cibles proposés apparaît. Cochez la case pour TEI P5 XML.
- En haut de l'écran un bouton Browse apparaît. Cliquez sur ceci, et naviguez jusqu'au fichier ubu.doc dans votre dossier Travaux.
- Cliquez sur le grand bouton Convert et patientez.
- Le site vous renvoie (après un bref délai) un fichier ubu.xml. Enregistrez-le dans votre dossier Travaux.

11 Structuration des pièces de théâtre

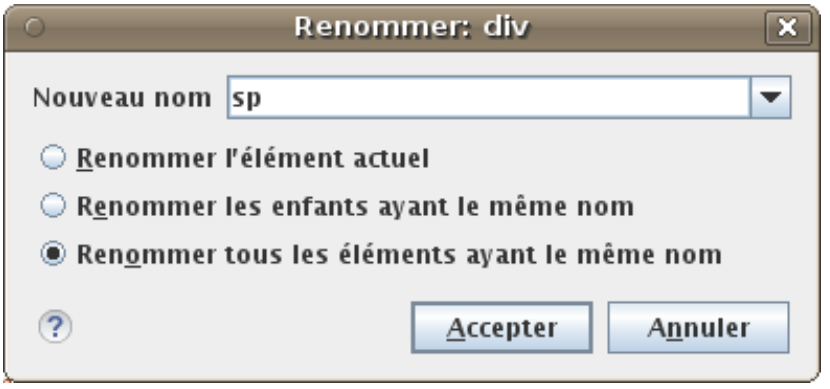
Une pièce de théâtre contient :

- des *didascalies* de plusieurs types... à baliser avec `<stage>` (en se servant de l'attribut type pour les distinguer au cas ou)
- des *énoncés* ou discours ... à baliser avec `<sp>` ('speech')
- des titres ou marques de *locuteur* ... à baliser avec `<speaker>`
- des paragraphes, ou des vers, balisés comme d'habitude.

12 Transformation des balises

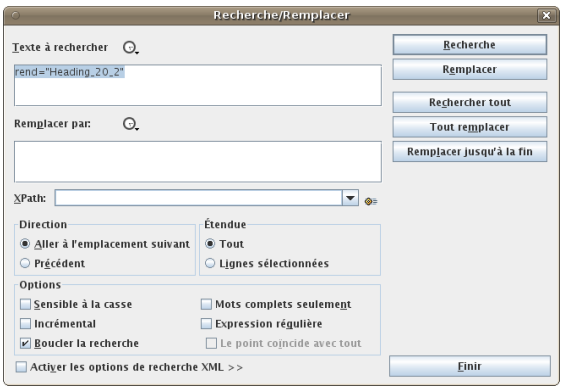
Voyons ce que le Garage aura fait de notre fichier Word. En tout cas, il est devenu un document TEI valide, même s'il est plein de mensonges...

- Lancez Oxygen, et ouvrez le fichier ubu.xml que vous venez de créer.
- On laisse passer pour le moment le TEI Header. Vous pouvez le compléter plus tard.
- D'abord notez que chaque énoncé est devenu un `<div>`, contenant un `<head>` et un `<p>`. Cela va beaucoup nous aider, et c'est une conséquence du fait que la version Word s'est servi des styles (chaque énoncé était précédé d'un objet style 'heading 2')
- Mettez le curseur d'abord sur une des balises `<div>`
- Dans le menu Document sélectionnez la commande XML-Refactoring, puis Renommer l'élément (ou tapez ALT-MAJ-R).
- Oxygen vous propose ce dialogue



- Dans le champs `Nouveau nom` entrez `sp`, le nom de la balise souhaitée, et cocher la case Renommer tous les éléments ayant le même nom.
- Cliquez sur Accepter et tous les `<div>` deviennent `<sp>`. (Le document n'est plus valide, mais on va rectifier cela tout à l'heure.)

- Faites de même pour les balises `<head>` : transformez les tous en `<speaker>`.
- Avant de rectifier la structuration du document, on va supprimer les `rend="Heading_20_2"` qui n'ont plus de sens. Sélectionnez cette chaîne de caractères n'importe où sur l'écran.
- Tapez CRL-F, ou sélectionnez Rechercher/Replacer sur le menu Recherche.



- Tapez `Tout remplacer`.

Enfin, regardez les trois morceaux de texte mal-balisés au début de la scène : le plus simple serait d'enlever leur balisage actuel et ensuite d'ajouter les balises qu'il nous faut.

- Mettez le curseur sur le texte 'Scène Vi'. Dans le menu Document, sélectionnez XML Refactoring, et ensuite Effacer les balises (ALT-MAJ-X). Répétez jusqu'a ce que toutes les balises qui l'entourent soient effacées.
- Même jeu pour les deux lignes suivantes : 'Le palais...' et 'LE ROI VENCESLAUS...'
- Maintenant, entourez tout le texte souligné en rouge, et balisez-le avec un `<head>`.
- Votre document redevient valide ! à vous maintenant de le rendre correcte :
  - diviser les trois parties du `<head>` au début de la scène.
  - les phrases balisées `<emph>` et l'un des `<p>` sont des didascalies : changez- les donc en `<stage>`.
  - il reste des valeurs de l'attribut rend à enlever
  - Dans ce `<body>` nous n'avons qu'une des scènes du texte d'Ubu. Il serait préférable (plus honnête !) d'indiquer cela en entourant le tout avec un `<div type="scene">`

S'il vous reste du temps, pensez à compléter le TEI Header. Quand vous aurez fini, n'oubliez pas d'enregistrer votre belle version TEI du fichier!