

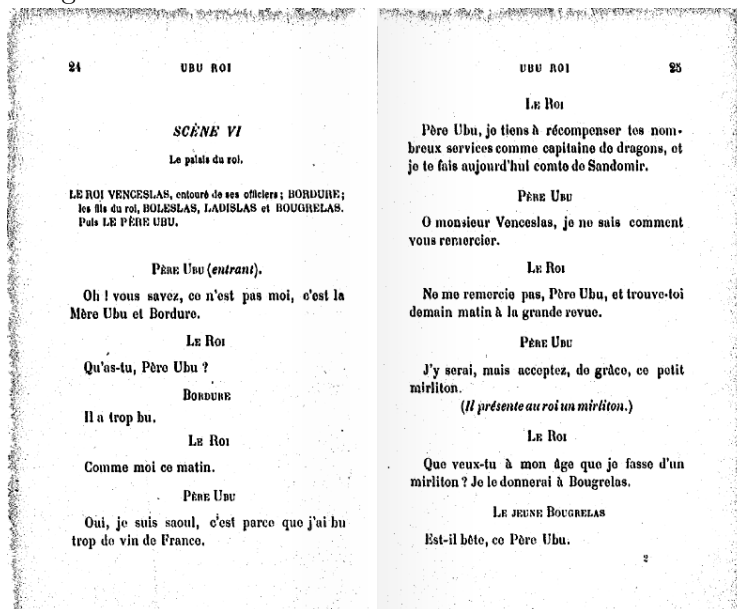
1 À la découverte d' Oxygen - 3

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

- l'encodage XML d'un document Word existant
- la représentation des structures aperçues dans un pièce de théâtre

2 Par ma chandelle verte...

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



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.

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 mirilton.

3 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 perdriions 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.

4 Structuration des pièces de théâtre

Une pièce de théâtre contient :

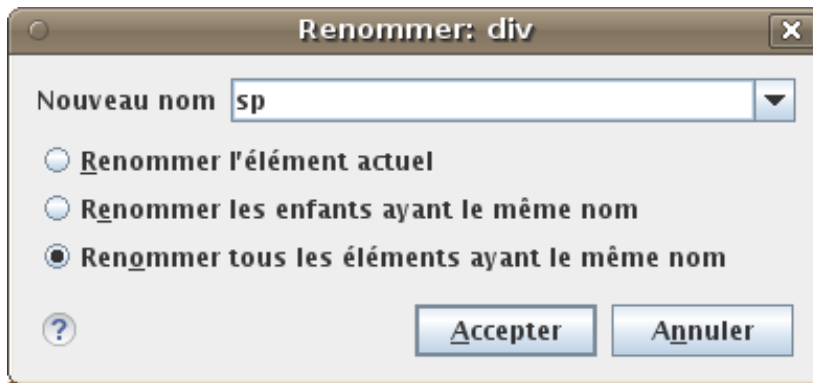
- des *didascalies* de plusieurs types... à baliser avec `<stage>` (en se servant de l'attribut `type` pour les distinguer au cas ou)
- des *énoncés* ou discours ... à baliser avec `<sp>` ('speech')
- des titres ou marques de *locuteur* ... à baliser avec `<speaker>`
- des paragraphes, ou des vers, balisés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

5 Transformation des balises

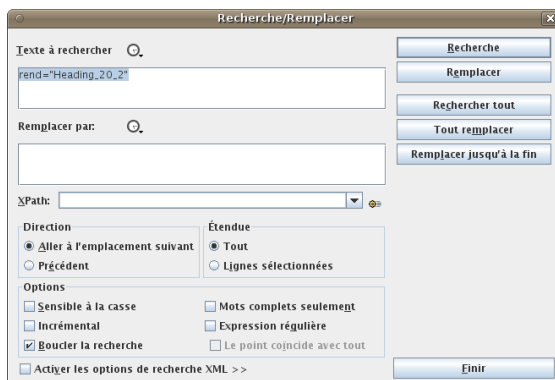
Voyons ce que le Garage aura fait de notre fichier Word. En tout cas, il est devenu un document TEI valide, même s'il est plein de mensonges...

- 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/Remplacer 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'à 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!