

À la découverte d'oXygen

Lou Burnard Consulting

2017-03-10

Dans ce premier exercice, nous allons nous servir d'oXygen pour :

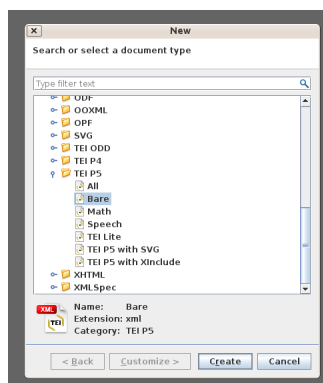
- la création d'un nouveau document XML
- le balisage progressif d'un document
- la validation continue d'un document
- l'affichage du document sans balise

Si vous êtes déjà familiers avec oXygen, vous pouvez directement passer au deuxième exercice (Importation d'un document bureautique).

1 Création d'un document nativement-numérique

La TEI ne sert pas uniquement à l'encodage des documents anciens, déjà existants sur papier ou pierre. Elle peut aussi servir à la création de nouveaux documents créés *ex nihilo*.

- Démarrez oXygen.
- Cliquez sur l'icône Nouveau en haut à gauche (ou sélectionnez Nouveau du menu Fichier, ou bien tapez `Ctrl+N`) pour ouvrir la boîte de dialogue Nouveau
- Choisissez Cadre des modèles, ensuite TEI P5, puis Bare, pour sélectionner un schéma TEI minimal.



- Cliquez sur le bouton Créer en bas. oXygen affiche un document TEI que vous allez compléter.

Le schéma TEI bare contient un minimum de balises, mais il permet quand-même de décrire un document TEI complet.

On est maintenant dans l'espace de travail typique d'oXygen. Par défaut, il y a plusieurs petites fenêtres, autour d'une grande zone blanche où nous allons éditer notre document. À gauche, par exemple, vous voyez dans la fenêtre Sommaire un aperçu de la structure du document.

Notez qu'un document TEI est composé de deux parties. En haut, vous avez les métadonnées, entourées par la balise `<?xml version="1.0" encoding="UTF-8" standalone="no" ?>`. En bas se trouve le texte, entouré par la balise `<body>`.



En bas de l'écran, vous voyez qu'il est possible de modifier l'affichage du document en cliquant sur les boutons *Texte*, *Grille*, ou *Auteur*. Nous, on n'a pas peur des balises, donc on va commencer à l'éditer en mode *Texte* !

D'abord, on va fournir les métadonnées essentielles à tout nouveau document : son titre, sa diffusion, et ses origines. oXygen vous propose des morceaux de texte qu'il faut changer puis compléter...

- Remplacez le mot 'Title' (entre les balises `<title>` et `</title>`) par un titre pour votre document. À vous de choisir ! Quelque chose comme 'À la découverte d'oXygen' serait suffisant.
- Remplacez les deux autres morceaux de texte proposés par des mentions plus appropriées. Par exemple, pour publication, on pourrait mettre 'Brouillon inédit' ou 'Distribué sur mon blog' ; pour , on vous propose 'Œuvre originale, guidée par une formation TEI'.

Par défaut, oXygen cherche à maintenir votre document dans un état valide. Qu'arrive-t-il si vous endommagez votre document ?

- Effacez quelques lettres à l'intérieur de n'importe quelle balise : par exemple changez `<title>` en `<titl>`.
- oXygen fait de son mieux : D'abord, il change la balise fermante correspondante (qui devient instantanément `</titl>`) : le document reste donc bien formé. Mais il n'est plus valide par rapport au schéma TEI bare (qui ne contient aucune balise 'publicat'). Par conséquent, les balises fautives sont soulignées en rouge ; il y a un message d'erreur en bas de l'écran ; et à droite il y a quelques taches rouges.
- Cliquez sur les taches rouges pour essayer de comprendre la manière dont une erreur peut en entraîner une autre.
- Cliquez sur (ou sélectionnez **Undo** du menu **Edit**) pour revenir dans un état de grâce... vous vous apercevrez que le petit carré rouge en haut à droite reprend une jolie couleur verte après correction de toutes les erreurs.

En effet, oXygen ne vous permettra d'ajouter que des balises prévues dans le schéma actuellement sélectionné pour le document.

- Mettez le curseur après la balise fermante et avant la balise fermante ;
- tapez et arrêtez un instant ;
- le carré vert repasse au rouge, parce que vous n'avez pas encore entré le nom de l'élément souhaité.
- oXygen vous propose une liste de balises disponibles à cet emplacement dans le document : notamment vous avez le droit d'insérer ici un élément `<div>` ou encore un `<comment>`, de terminer le `<div>`, d'insérer un commentaire XML etc.

-
- Tapez sur ENTRÉE pour accepter la première proposition. oXygen insère dans votre document les balises `<?xml version="1.0" encoding="UTF-8" standalone="no" ?>` et `<?xml-stylesheet href="oxygen.xsl" type="text/xsl" ?>`.
 - Le curseur reste à l'intérieur de la balise ouvrante, au cas où vous désireriez spécifier ses attributs. Faites bouger le curseur d'un caractère vers la droite et tapez votre nom, comme auteur du fichier.

2 Ajout de texte au document

Enfin, nous sommes prêts... commencez la création de votre document. À vous de décider de son contenu et de son balisage.

D'abord, quels composants proposeriez-vous pour votre document ? Un document contenant un seul paragraphe n'est guère probable ! Avec le schéma *bare* vous avez le droit de signaler :

- des titres... à baliser avec la balise `<h1>`
- des chapitres... à baliser avec la balise `<h2>`
- des listes... à baliser avec la balise ``
- des paragraphes... à baliser avec la balise `<p>`

Allons-y !

Dans le reste de ce petit tutoriel, nous allons vous montrer quelques manipulations utiles offertes par ce logiciel.

D'abord, une question importante : Pour le contenu, OK, il faut le créer — mais est-ce qu'il faut toujours taper soi-même les balises ? Non : oXygen est là pour vous aider. Vous avez déjà vu comment il peut vous proposer la liste des balises disponibles à un certain emplacement du document. Il peut aussi vous suggérer comment baliser un morceau de texte sélectionné.

- Avec la souris, sélectionnez le contenu du `<p>`, i.e. la séquence `<p>some text here</p>`.
- Tapez `<` (ou sélectionnez XML Refactoring et ensuite Entourer des balises dans le menu Document).
- oXygen vous propose toutes les balises qui pourraient éventuellement être insérées à cet endroit précis du document : dans ce cas, il est conseillé de sélectionner `<p>`. Puis, cliquez sur Accepter
- L'élément `<p>` va contenir le premier chapitre de votre document. Insérez donc un titre au début de ce `<p>`, en vous servant de la balise `<h1>`
- Il serait aussi utile de numéroté vos chapitres d'une manière explicite : pour cela, on vous propose l'attribut `@n` qui est disponible sur n'importe quel élément TEI, y compris `<h1>`.
- Mettez le curseur à l'intérieur de la balise `<h1>` juste avant le `>` et tapez un seul espace. oXygen vous propose un menu des attributs disponibles.
- Tapez sur ENTRÉE pour sélectionner `@n`. oXygen vous informe qu'il faut dans ce cas ajouter une valeur, et le curseur reste entre les guillemets.
- Tapez (par exemple) `1`, et votre document redeviendra valide.

À vous d'ajouter maintenant du texte à ce paragraphe : un peu plus intéressant que 'some text here' de préférence ! Décrivez-nous vos premières réactions avec oXygen ; partagez vos idées sur la situation politique actuelle ; écrivez ce que vous voulez.

Sans doute aurez-vous besoin de plus d'un seul paragraphe. Comment faire pour terminer celui-ci et commencer un nouveau ? Bien-sûr, on peut simplement taper les balises mais il y a une plus simple manière de faire.

Supposons que vous avez maintenant quelques centaines de mots dans votre seul paragraphe, et que vous voulez le diviser en plusieurs.

- Mettez le curseur au point de chaque division souhaitée
- Tapez ou sélectionnez XML Refactoring et ensuite Élément de division dans le menu Document.

Vous pouvez faire pareil pour diviser le chapitre unique, mais attention à choisir un endroit valide pour effectuer la division : *entre* deux éléments , et non pas dans l'enceinte d'un .

Dans un de vos chapitres, n'omettez pas d'ajouter une liste, balisée avec un élément , contenant une séquence de éléments, précédée d'un facultatif.

3 L'affichage du texte balisé

Est-ce que la quantité de balises dans votre document commence à vous inquiéter un peu ?

On peut mettre en ordre le balisage que l'on a fait, en cliquant sur le bouton d'indentation, ou en tapant , ou en sélectionnant Source->Indenter le document sur le menu Document. Mais peut être préféreriez-vous voir disparaître les balises ?

- Au fond de la fenêtre vous voyez trois onglets : Texte, Grille, et Auteur. Sélectionnez le dernier. L'affichage de votre document change, et un nouveau menu TEI P5 devient disponible. Sur ce menu, sélectionnez Mode d'affichage des balises et regardez les possibilités offertes par ces options.
- L'option Toutes les balises affiche les balises en icônes ; l'option Aucune balise les fait disparaître totalement.
- Notez en plus que votre document est maintenant stylé ; il y a des changements de police ; les retours de ligne sont traités comme des blancs ; etc.
- Dans le mode auteur, on peut diviser un élément simplement en tapant ENTRÉE deux fois. Vous pouvez essayer.

Plus tard, nous verrons comment vous pourrez contrôler l'affichage de votre document avec une *feuille de style*.

4 Importation d'un document bureautique

La plupart des documents numériques que vous rencontrerez dans le monde actuel paraissent indemnes de toute balise XML. Par exemple, les documents préparés à l'aide des outils bureautiques tels Microsoft Word ou Open Office se présentent en apparence comme du texte brut.

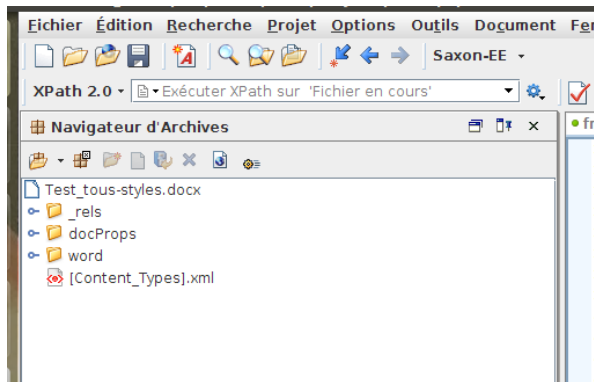
En réalité, ce n'est qu'une apparence : derrière tout cela il existe un format XML, que l'on peut manipuler, et donc qu'on peut transformer dans un format XML TEI.

Dans votre dossier `exercices` vous trouverez un fichier nommé `acteRoyal-s.docx` et un autre `acteRoyal.jpg` : Prenez-en connaissance.

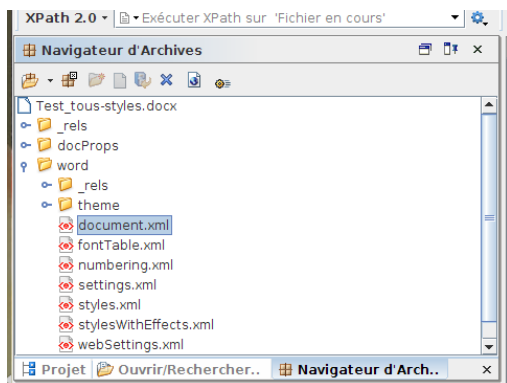
Il s'agit d'un document numérisé en mode image qui a ensuite été rapidement retranscrit avec Microsoft Word. On va se servir d'oXygen pour transformer cette transcription en un document XML-TEI.


Un fichier Microsoft Word au format docx est une archive compressée qui contient plusieurs autres fichiers, y compris des fichiers XML qu'on peut gérer et transformer avec oXygen comme n'importe quel autre fichier XML. La conversion prend en compte les styles du document existants dans le fichier, et donc peut être personnalisée.

- Ouvrir oXygen. Sélectionnez Ouvrir dans le menu Fichier (ou tapez `Ctrl+O`, ou bien cliquez sur l'icône en forme de dossier) et sélectionnez le fichier `acteRoyal-s.docx` qui se trouve dans le répertoire `exercices/`.
- Une fenêtre intitulée **Navigateur d'Archives** s'ouvre à gauche de l'écran principal. La structure de l'archive docx y est affichée.

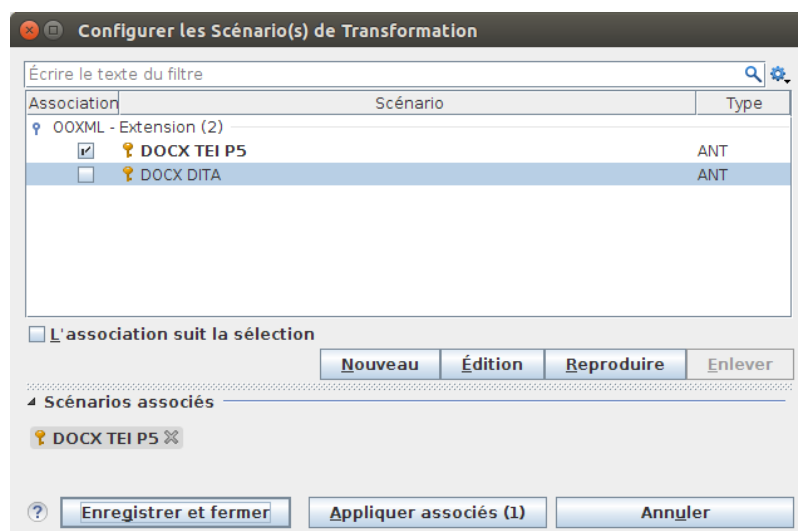



- Cliquez sur la petite clef bleue à côté du dossier `word` pour voir ce qu'il contient



- Sélectionnez le fichier `document.xml` et double-cliquez pour l'ouvrir (cela peut prendre un peu de temps si le fichier est volumineux)
- Voilà : c'est un document XML, plein de balises, pour la plupart appartenant à un espace de nom défini par Microsoft. Convertissons-le !
- Avec ce fichier `document.xml` ouvert dans l'écran principal, sélectionnez **Transformation -> Configurer les Scénario(s) de Transformation** du menu Document. Ou tapez `Ctrl+Alt+T`. Ou cliquez sur la petite icône en forme de clef à molette (celui-ci : )
- oXygen vous présente les options de transformation disponibles pour ce document, parmi lesquelles devrait figurer une transformation nommée **DOCX TEI P5**.

4 IMPORTATION D'UN DOCUMENT BUREAUTIQUE



- Sélectionnez la case à cocher correspondante et cliquez sur le bouton **Appliquer associés** au dessous. Une nouvelle fenêtre s'ouvre en bas pour l'affichage des messages concernant le processus de transformation : vous pouvez les ignorer sans problème.
- Dans l'écran principal, un document TEI-conforme s'affiche. Sélectionnez **Source -> Formater et indenter** du menu **Document** menu, ou tapez ou cliquez sur le bouton **Indent** (celui-ci : ) pour regarder sa structure
- L'attribut *@rend* indique le nom du style Word utilisé dans cette portion du document original. Une transformation XSLT plus complexe pourrait donc s'en servir pour effectuer une conversion plus intelligente.

La question qui s'impose : quelles balises devraient être disponibles pour une représentation efficace des aspects les plus signifiants du document ?