

# Oxygen: tutoriel

Lou Burnard

juin 2011

## 1 À la découverte d'Oxygen - 1

Dans la première partie de ce petit tutoriel, 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

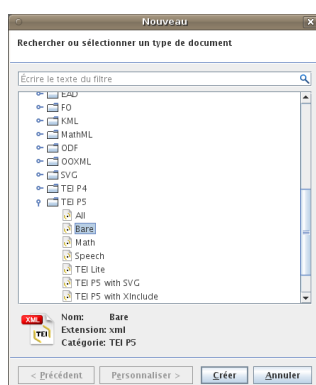
Si une version récente du logiciel Oxygen n'est pas encore installé sur votre poste, il faut l'installer avant d'aborder ce tutoriel. Vous pouvez le télécharger du site Oxygen à [http://www.oxygenxml.com/download\\_oxygenxml\\_editor.html](http://www.oxygenxml.com/download_oxygenxml_editor.html) et puis l'installer avec une licence d'essai, gratuit et valable 30 jours.

### 1.1 Le document né-numérique

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

### 1.2 Création d'un document

- Démarrer 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
- Choisissez Modèles du framework, ensuite TEI P5, puis Bare, pour sélectionner un schéma basique TEI.

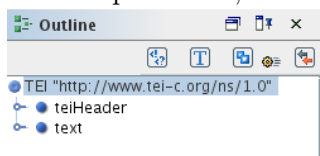


- Cliquez sur le bouton Créer en bas. Oxygen affiche un document TEI minimal que vous allez compléter.

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

On est maintenant dans l'espace de travail typique d'Oxygen. Par défaut, il y a plusieurs petits écrans, autour d'un grand espace blanc où nous allons éditer notre document. À gauche,

par exemple, vous voyez dans la fenêtre **Sommaire** un aperçu de la structure du document. (Si ce n'est pas le cas, sélectionnez **Fenêtre -> Afficher la vue**, et ensuite **Sommaire**)



Vous voyez qu'un document TEI est divisé en deux parties. En haut, vous avez les métadonnées, entourées de la balise `<teiHeader>`. En bas, se trouve le texte, entouré de la balise `<text>`.



En bas de l'écran, vous voyez qu'il est possible 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 l'éditer en mode *Texte*.

D'abord, il faut fournir les métadonnées essentielles à tout nouveau document : son titre, sa distribution, 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' suffirait.
- 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' selon le cas ; pour `<sourceDesc>`, pour `<sourceDesc>`, on vous propose 'Œuvre originale, guidé par un tutoriel TEI à MISAT 2016'.

Oxygen par défaut cherchera à maintenir votre document dans un état de validité. Qu'est-ce qui arrive si vous l'endommagez ?

- Effacez quelques lettres à l'intérieur de n'importe quelle balise : par exemple changez `<publicationStmnt>` en `<publicat>`.
- Oxygen fait de son mieux. D'abord, il change la balise fermante correspondante (qui devient instantanément `</publicat>`): le document reste donc bien formé. Mais il n'est plus valide : 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 CTRL-Z (ou sélectionnez **Undo** du menu **Edit**) pour revenir dans un état de grâce... vous verrez que le petit carreau rouge en haut à droite devient vert après correction de toutes les erreurs.

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

- Mettez le curseur à la fin de la 7ème ou au début de la 8ème ligne, entre `</title>` et `</titleSmt>`;
- tapez `<` et arrêtez un instant ;
- le carreau vert redevient 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 `<author>` ou encore un `<title>`, de terminer le `<titleSmt>`, d'insérer un commentaire XML etc.
- Tapez sur ENTRER pour accepter la première proposition. Oxygen insert dans votre document les balises `<author>` et `</author>`.
- Le curseur reste à l'intérieur de la balise ouvrante, au cas où vous désireriez spécifier ses attributs. Faire bouger le curseur d'un caractère à droite et tapez votre nom, comme auteur du fichier.

### 1.3 Ajout de texte au document

Enfin, nous sommes prêts pour commencer la création de notre 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 divisions ... à baliser avec la balise `<div>`
- des titres ... à baliser avec la balise `<head>`
- des chapitres ... à baliser avec la balise `<lg>`
- des listes ... à baliser avec la balise `<list>`
- 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 proposer comment baliser un morceau de texte sélectionné.

- Avec la souris, sélectionnez le contenu du `<body>`, i.e. la séquence `<p>Some text here</p>`.
- Tapez CTRL-E (ou dans le menu Document sélectionnez Balisage et ensuite Encadrer avec des balises ).
- 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 `<div>`. Puis, cliquez sur OK

- L'élément **<div>** va contenir le premier chapitre de votre document. Nous allons insérer donc un titre au début de ce **<div>**, en nous servant de la balise **<head>**. Vous connaissez déjà la manip: tapez un chevron au bon endroit et sélectionnez le nom de la balise qu'il vous faut!
- Il serait aussi utile de numéroté vos chapitres d'une manière explicite : pour cela, nous vous proposons l'attribut *@n* qui est disponible sur n'importe quel élément TEI, y compris **<div>**.
- Mettez le curseur à l'intérieur de la balise **<div>** juste avant le **>** et tapez un seul espace. Oxygen vous propose un menu des attributs disponibles.
- Tapez sur RETOUR pour sélectionnez *@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.

Votre document devrait maintenant ressembler à ceci :

```
<text>
  <body>
    <div n="1">
      <head>Chapitre premiere</head>
      <p>Some text here</p>
    </div>
  </body>
```

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 **</p><p>** mais il y a une autre manière plus simple.

### 1.4 Division des éléments

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 ALT-MAJ-D ou bien dans le menu Document sélectionnez Balisage et ensuite Scinder l'élément

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

Dans un de vos chapitres, n'omettez pas d'ajouter une liste, balisée avec un élément **<list>**, contenant une séquence de **<item>** éléments, précédée d'un **<head>** facultatif.

### 1.5 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 CTRL+MAJ+P, 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 ses options.

- L'option **Toutes les balises** affiche les balises en icones; l'option **Aucune balises** les fait disparaître totalement.
- Notez en plus que votre document est maintenant stylé ; il y a des changements de police ; les retours de chariots sont traités comme des blancs ; etc.
- Dans mode auteur, on peut diviser un élément simplement en tapant RETOUR deux fois. Expérimentez cela. c

## 2 À la découverte d' Oxygen - 2

Vous avez sans doute noté que le schéma TEI Bare dont nous nous sommes servis dispose d'un nombre très limité de balises: à partir d'un `<div>` on n'a droit qu'à `<p>` et `<list>` et leurs composants `<label>`, `<title>`, et `<author>` . Ce n'est guère satisfaisant: on pourrait souhaiter par exemple baliser les noms propres avec `<name>`, ou les parties mises en evidences avec `<hi>`. On pourrait souhaiter enrichir le document avec de metadonnées plus exactes sur les personnes ou les lieux, ou avec des analyses linguistiques de chaque mot. Pour tout cela, il nous faut un schéma plus complet, plus adapté à notre projet.

Dans cette deuxième partie du tutoriel, nous allons revenir sur cette question d'un schéma TEI

- l'intérêt de limiter les balises, les attributs, et les valeurs d'attribut mises à disposition
- création d'un schéma personnalisé en utilisant Roma
- aperçu du format ODD

### 2.1 Création d'un document maximal

- Démarrez oXygen.
- Cliquez sur l'icône **Nouveau** en haut à gauche (ou sélectionnez **Nouveau** dans le menu Fichier, ou tapez CTRL-N) pour ouvrir la boîte de dialogue Nouveau
- Oxygen affiche une longue liste des types de document connus. Regardez sous **Cadre des modèles**, ensuite **TEI P5**, et puis cliquez sur **All**, pour sélectionner un schéma TEI maximal.
- Cliquez sur le bouton **Créer** en bas. oXygen affiche un document TEI.
- Enregistrez ce fichier dans votre dossier **Travaux** sous le nom **myTest.xml**.

### 2.2 All n'est pas la solution idéale

Comme vous l'avez déjà vu, oXygen propose éléments et attributs TEI automatiquement.

- Dans le document que vous venez de créer, mettez le curseur à l'intérieur de l'élément `<p>`.
- Tapez `<`. oXygen vous offre un menu de tous les éléments disponibles à ce point.
- Passez en revue la liste des noms d'élément : un "pop-up" apparaît pour expliquer la fonction de chacun. Regardez par exemple `<address>`, `<camera>`, `<incident>`, `<metamark>`, ou `<notatedMusic>`.
- Tapez ESC pour sortir du menu ; supprimez le '`<`' que vous venez d'ajouter.
- On a beaucoup de choix... mais dans un projet particulier il n'est guère probable qu'on ait besoin de tous ces éléments. Avec autant de possibilités, la probabilité d'introduire des incohérences ou d'effectuer un mauvais choix est fortement augmentée. Cela complexifie également le traitement des documents XML que nous allons créer.

### 2.3 Création d'un schéma avec Roma

Le logiciel Roma vous permet de créer votre propre schéma TEI. Notre objectif sera de créer un schéma pour traiter (uniquement) les imprimés, en utilisant les balises que vous avez déjà découvertes.

- Avec votre navigateur rendez-vous sur <http://www.tei-c.org/Roma/>. La page vous propose quatre options:
  1. *Build up* : on crée une personnalisation en ajoutant petit à petit ce qu'on souhaite avoir en partant d'une base minimale
  2. *Reduce* : on fait l'inverse : c'est-à-dire qu'on crée une personnalisation en retirant petit à petit ce qu'on ne souhaite pas avoir en partant d'une base maximale
  3. *Create* : on part d'un template déjà existant
  4. *Use* : on se sert d'une personnalisation déjà proposée par la TEI
  5. *Upload* : on retravaille une personnalisation qu'on a déjà enregistrée.
- Dans cet exercice, nous choisissons la troisième solution. Choisissez **Create** et sélectionnez **TEI Bare**, puis cliquez sur le bouton rouge **start**.
- Pour régler les paramètres on vous propose :
  - *Title* : Changez 'TEI absolutely bare' en (par exemple) "TEI pour les corpus".
  - *Filename* : Change 'tei\_bare' en (par exemple) **teiCorpus** (c'est un identifiant, et donc ne devrait pas inclure de blancs)
  - *Language* : Vous pouvez laisser en anglais, mais ce tutoriel part du principe que vous préférez travailler en français. Cochez donc le bouton devant le mot "français".
  - *Author name* : Entrez votre propre nom (ou celui de votre avatar si vous préférez)
  - Modifiez la description si vous le souhaitez.
- Cliquez sur le bouton rouge **Save** en bas de page. Vos changements ont été pris en compte : l'interface est maintenant en français, et en haut à droite, le nom de la personnalisation sur laquelle vous travaillez est affiché.
- Roma est un logiciel international (plus ou moins). Sélectionnez l'onglet **Langage**, et cliquez sur le bouton **Français** pour assurer que la francisation de cet exercice soit maximale !
- Encore une fois, cliquez **Save** en bas de page.

### 2.4 Sélection de modules

Un *module* est un regroupement d'éléments TEI. Tout élément TEI est déclaré dans un module spécifique. Par exemple, pour encoder un dictionnaire vous aurez besoin des éléments spécialisés pour les dicos. Mais ces éléments ne sont pas forcément intéressants pour l'encodage d'autres types de documents.

Dans Roma,

- Cliquez sur l'onglet **Modules** pour vérifier quels modules constitueront votre schéma
- Nous avons commencé avec le schéma 'bare'. Donc, la liste des modules sélectionnés, affichée à droite, ne contient que trois modules.
- Supposons que nous souhaitons disposer des balises pour traiter les figures et les tableaux. Nous aurons besoin donc d'ajouter un module supplémentaire qui s'appelle **figures**

- Cliquez sur le mot **ajouter** devant la mention du module **figures** dans la liste à gauche. Ce module va apparaître tout de suite dans la liste à droite.
- Vous avez maintenant un schéma construit à partir des modules : **tei** (Roma ne vous permet pas d'enlever ce module infrastructurel), **core**, **header**, **figures** et **textstructure**.

## 2.5 Inclusion et exclusion des éléments

Dans ce petit ensemble de modules, plusieurs éléments sont supprimés; et peut être il y a d'autres qui n'ont aucun intérêt pour notre projet.

- Cliquez sur le mot *core* à droite (nota : non pas sur le mot *supprimer* mais sur le nom du module). Une liste de tous les éléments du module **core** s'affiche.
- Chaque ligne de cette table contient :
  - le nom canonique d'un élément
  - une indication de son Inclusion ou Exclusion
  - le nom de cet élément dans le schéma actuel (normalement c'est le nom canonique i.e. anglais)
  - un point d'interrogation fournissant un lien vers la page de référence pour cet élément
  - une description brève des fonctions de cet élément
  - un lien qui permet de modifier les attributs de cet élément
- Cette interface vous permet d'explorer la fonction et l'usage de tous les éléments TEI. Il vous permet également de sélectionner ceux que vous souhaitez intégrer dans votre schéma.
- Vous pouvez Inclure ou Exclure tous les éléments d'un module. Cliquez sur le mot **Exclure** dans le titre de la table, et tous les éléments sont exclus par défaut. Pareille pour le mot **Inclure** bien sur. Cliquer le bouton à côté du nom d'un élément pour changer son in/ex-clusion!
- A vous de décider quels éléments vous souhaitez inclure. Pour commencer, vous êtes invité à ajouter deux éléments **<name>** et **<graphic>**.
- Ensuite considérez les éléments ajoutés dans votre schéma par le module **figures** : nous vous invitons d'en supprimer au moins **<annotatedMusic>**
- N'oubliez pas de cliquer sur le bouton rouge **Save** en bas de page !

## 2.6 Création de schéma

- Cliquez sur l'onglet **Schéma**. Vous pouvez choisir entre plusieurs langages de schéma : la TEI se définit indépendamment de toute langue, afin de pouvoir générer tous ces formats.
- Nous vous conseillons de générer un schéma ou bien en Relax NG Compact Syntax, ou en Relax NG XML Syntax.
- Cliquez sur **Générer** et votre navigateur vous envoie un fichier de schéma. Enregistrer ce fichier dans votre dossier Travaux.
- **Ne fermez pas votre navigateur : vous en aurez encore besoin.**

### 2.7 Association de schéma et fichier en oXygen

Jusqu'à présent, oXygen associait le schéma `tei_all` avec votre fichier `myTest.xml`.

- Revenez dans oXygen et rouvrez votre fichier
- Dans le menu Document ouvrir le menu Schéma et sélectionnez Associer un schéma. (Notez l'icône bleue et rouge pour cette action qui devrait également apparaître sur la barre à outils)
- Dans la boîte de dialogue qui s'affiche, cliquez sur la petite icône de dossier à côté de URL pour naviguer vers le fichier RNG que vous venez de créer avec Roma. Cliquez sur OK.
- Dans votre fichier XML-TEI, vous devrez maintenant voir une ligne qui ressemble à ceci:

```
<?xml-model
compact-syntax"                                href="teiCorpus.rnc" type="application/relaxng-
ns="http://relaxng.org/ns/structure/1.0"?>
```

. S'il y a d'autres lignes du même type, enlevez-les pour être sûr !

### 2.8 Confirmation

Vous vous souvenez de ces éléments bizarres `<address>`, `<camera>`, `<musicNotation>` etc. qui étaient disponibles au sein d'un paragraphe ?

- Mettez le curseur après une balise ouvrante `<p>`, et tapez '`<`' pour voir de nouveau la liste des éléments disponibles.
- Tous les éléments que vous avez supprimés ont disparu ? C'est bien ! Sinon, peut-être que vous avez oublié de cliquer sur le bouton 'Save' après les avoir inclus/exclus... essayez encore !

### 2.9 Comment limiter les valeurs de l'attribut `@type` de l'élément `<name>`

Nous avons vu qu'il est très utile de supprimer tel ou tel élément de notre schéma. Que peut-on faire pour limiter les valeurs légales de ses attributs ? (Notez qu'en général les attributs TEI sont définis d'une manière très permissive : vous pouvez taper n'importe quoi). Supposons que nous désirons limiter les valeurs possibles pour l'attribut `@type` sur `<name>`.

- Revenez dans Roma
- Cliquez sur l'onglet Modules
- Cliquez sur le module `textstructure`
- Trouvez le rang où est défini l'élément `<name>` et cliquez sur **Changer les Attributs** tout à droite.
- Une table des attributs disponibles pour cet élément s'affiche. Vous pouvez inclure ou exclure des attributs si cela vous intéresse. Pour le moment essayons quelque chose d'autre.
- Descendez jusqu'à `@type` et cliquez sur l'attribut. Une page s'affiche, qui vous permet de vérifier plusieurs options pour cet attribut :
  - *Facultatif* ? Est-ce qu'il faut fournir une valeur pour cet attribut ou est-il facultatif ? Supposons que nous désirons un schéma dans lequel la présence d'une valeur de `@type` sur `<div>` soit obligatoire : dans ce cas, sélectionnez 'no'.



- *Contenu* : Combien de valeurs (une ou plusieurs) sont permises pour cet attribut, et de quelle type (texte, numéro, date...) ? Choisissons une seule valeur de type 'text'.
- *Valeur par défaut* : nous permet de proposer une valeur par défaut pour l'attribut, dans le cas où aucune valeur n'est fournie dans le document. Laissez cela vide, puisque nous avons déjà décidé de rendre obligatoire la présence d'une valeur.
- *Liste fermée* : nous permet de préciser si la liste de valeurs proposées est une liste fixe impossible à augmenter ou non. Dans notre schéma, c'est une liste fermée, donc répondez 'Yes' !
- *Liste de valeurs* : nous permet de spécifier les valeurs permises pour cet attribut dans notre schéma. Les valeurs sont proposées sous la forme d'une seule ligne, chaque valeur séparée par une virgule. Entrez donc la séquence suivante (*sans blancs*) **personne,lieu,societe,autre**.
- *Description* : nous permet de modifier la description de cet attribut. Vous pouvez la changer comme vous le souhaitez : par exemple, changer "n'importe quel système ou typologie de classification approprié" en "le système de classement proposé pour notre projet".

- Cliquez sur le bouton **Save** en bas de la page !

## 2.10 Essayons de nouveau...

Est-ce que nous avons réussi? Vous connaissez maintenant la démarche... :

- Cliquez sur l'onglet Schéma
- Choisissez l'un des formats de schéma
- Cliquez sur Générer
- Trouvez le fichier généré et téléchargé; copiez-le sur l'ancienne version pour la remplacer
- **Ne fermez pas votre Navigateur Internet !**
- Revenez dans votre fichier myTest.xml dans oXygen; choisir 'Document -> Valider -> Validate
- Si tout va bien, votre document reste valide : il y a un petit carré vert.
- Maintenant il faut ajouter une **<name>**. Comme vous avez fait avant, sélectionnez une denomination quelconque dans ce document, et entourez-le d'une balise **<name>** en tapant CTRL-E.
- Votre document n'est plus valide : il y a un carré rouge en haut ! En bas vous voyez le message **Element 'name' missing required attribute 'type'**
- Mettez le curseur à l'intérieur de la balise ouvrante de la **<div>**, juste avant le **>** et tapez un blanc. oXygen propose une liste d'attributs disponibles. Scrollez-la jusqu'à *@type* et notez que cet attribut est affiché en gras, pour signaler que sa présence est obligatoire.
- Sélectionnez *@type* et notez que oXygen vous propose la liste des valeurs que nous avons prédéfinies dans notre schéma.
- Choisissez l'une de ces valeurs, p.ex. **personne**, et votre document devient valide, avec un petit carré vert.

#### 2.11 Génération de documentation

Tout projet aura besoin de sa documentation interne, qui ne sera pas forcément optimale si elle est en RelaxNG ! Roma vous permet de générer automatiquement des spécifications compréhensibles, ressemblant à la doc de référence de la TEI.

- Retournez dans votre navigateur et cliquez sur l'onglet **Documentation**.
- Choisissez **HTML web page** et cliquez sur **Générer**.
- Après un bref délai, votre navigateur va recevoir un fichier HTML que vous pouvez lire avec le navigateur. Au début il y a une table des matières, avec un lien pour chaque élément de votre schéma. Scrollez jusqu'au lien pour **<div>** et cliquez dessus.
- Notez que la description de son attribut *@type* est modifiée selon vos propositions. (Mais notez aussi que d'autres détails, notamment les exemples d'usage, ne sont pas modifiés) .
- Si vous le préférez, vous pouvez également générer ce document en format PDF.

#### 2.12 Enregistrement de la personnalisation

Nous espérons que vous aurez maintenant une idée des capacités de ce système de personnalisation. Bien sûr, vous ne pouvez pas définir tout ce qui est le mieux pour votre projet en une seule session. Vous avez donc besoin d'une façon d'enregistrer la spécification, pour revenir dessus plus tard.

- Revenez dans votre navigateur et cliquez sur l'onglet **Enregistrer**.
- Votre navigateur vous envoie un fichier XML-TEI, avec le même nom que votre schéma, donc **teiCorpus.xml**. Enregistrez ce fichier dans votre dossier **Travaux**.
- Ouvrez ce fichier XML avec oXygen et lisez-le (mais avec les balises : en mode Auteur il y a des détails significatifs qui sont cachés !).
- D'abord notez que votre personnalisation hérite des propositions relatives au schéma minimal duquel vous l'avez dérivé. Vous pouvez bien sûr modifier ces propositions, en décrivant les objets de votre projet, vos procédures éventuelles de définition de schéma, vos règles maison de balisage, etc. etc. – juste comme un manuel d'encodage classique.
- Ensuite, regardez l'élément **<schemaSpec>**. Cet élément contient l'essentiel de votre personnalisation. Notez qu'il compte un élément **<moduleRef>** pour chacun des modules que vous avez sélectionnés, et que le module **core** est complété avec une liste des éléments que vous avez exclus de ce module.
- Regardez la spécification pour **<div>** (dans un élément **<elementSpec>**) pour voir comment nous l'avons modifié.
- Ce fichier est un document XML-TEI comme tous les autres. Vous pouvez donc faire des modifications, par exemple au début, pour améliorer votre manuel d'encodage. Pour recharger votre document dans Roma, il faut choisir la quatrième option 'Open existing customisation' sur la page d'accueil de Roma. S'il vous reste du temps, expérimentez cela !

### 3 À la découverte d' Oxygen - 3

Dans cette troisième partie du tutoriel, nous allons approfondir votre expérience avec Oxygen pour :

- l'encodage XML d'un document Word ou ODT existant
- le transformation d'un fichier XML dans un autre

### 3.1 Quelques documents historiques

Vous trouverez dans votre dossier Travaux un fichier nommé `petain-sample.docx`. Il contient trois communications du Maréchal Pétain, transcrits et annotés par notre collègue D. Mayaffre, ensuite édité dans un *Lexicometrica* de 2005.

Ce fichier contient :

- trois transcriptions distinctes, chacune datée
- des notices, qui figurent comme notes de bas de pages
- des paragraphes
- des titres

Pour ceux ou celles qui ne disposent pas de Microsoft Word, il y a aussi une version ODT du même fichier.

### 3.2 Transformation d'un fichier bureautique en TEI XML

Quoi faire pour transformer ce joli fichier en XML et travailler la-dessus avec Oxygen?

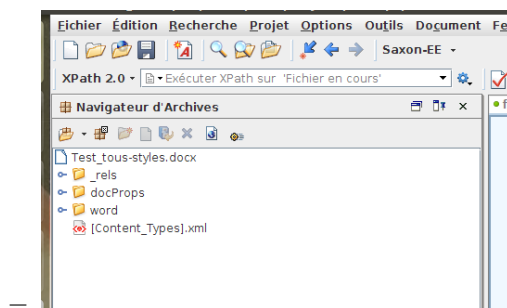
Nous pourrions l'exporter au format « plein text ». Nous pourrions aussi faire un copier-coller. Mais dans chaque cas, nous risquons la perte du formatage qui distingue (par exemple) les annotations en bas de page du texte annoté.

Si le fichier est bien stylé, on peut le transformer en XML, sans perte d'information.

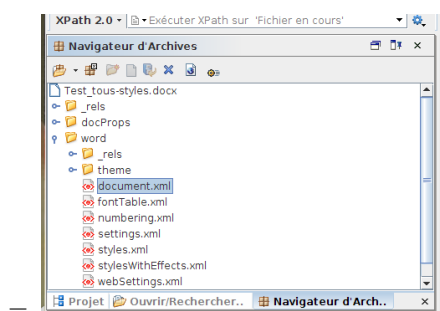
Un fichier `docx` ou un fichier `odt` est en effet une archive contenant plusieurs fichiers, notamment des fichiers XML susceptibles de manipulation et de conversion comme tout autre fichier XML. `oxyGen` est donc capable de le traiter, d'autant plus parce que ce logiciel est doté de plusieurs feuilles de style XSLT prédéfinies pour cette conversion en TEI XML. Ces feuilles font partie du 'TEI framework', qu'il faut installer dans votre copie d'`oxyGen` (mais qui est fournie comme un 'AddOn' par défaut dans les versions récentes)

(Ce TEI Oxygen framework peut être configuré à se mettre à jour automatiquement : voir un autre tuto pour vous renseigner sur cette démarche).

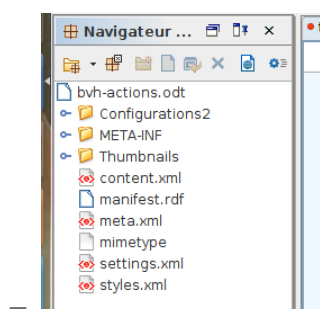
- Dans `oxyGen`, utilisez la boîte de dialogue habituelle Ouvrir pour sélectionner le fichier `petain-sample.docx` (ou `petain-sample.odt`) .
- Une fenêtre intitulée **Navigateur d'archives** s'ouvre à gauche de l'écran principal. Elle présente la structure de fichiers de l'archive `docx` ou `odt`, qui est légèrement différente dans les deux cas.
- **Microsoft Word**



- Cliquez sur la clef bleu à côté du répertoire appelé `word` pour voir le contenu de ce répertoire

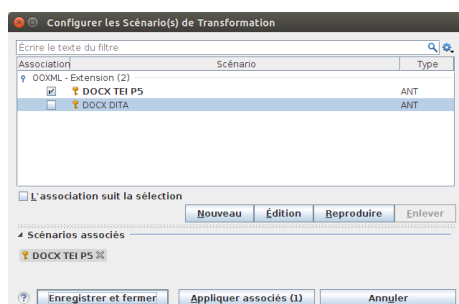



- Sélectionnez le fichier dénommé **document.xml** et double-cliquez pour l'ouvrir
- Open Office**



- Sélectionnez le fichier dénommé **content.xml** et double-cliquez pour l'ouvrir
- Comme vous le constatez, c'est un document XML, plein de balises XML utilisant plusieurs espaces de nom.

- Avec ce fichier (ou bien **document.xml**, ou bien **content.xml**) ouvert dans votre fenêtre d'édition principale, sélectionnez **Transformation -> Configurer Scénario(s) de Transformation** depuis le menu Document. Ou bien tapez **CTRL-SHIFT-C**. Ou cliquez sur l'icône avec la petite clef-à-molette
  - \* Pour la conversion par défaut d'un document DOCX vers TEI, cochez la case **DOCX TEI P5**
  - \* Pour la conversion par défaut d'un document ODT vers TEI, cochez la case **ODT TEI P5**
- et appuyez sur le bouton **Appliquer Associé(s)**



- Une nouvelle fenetre s'ouvre en bas de l'écran pour l'affichage des messages de traitement pendant la conversion; vous n'êtes pas obligé de les regarder...
- Dans la fenêtre d'édition principale, un document TEI-XML s'affiche (si tout va bien). Du menu **Document** sélectionnez **Source > Formater et Indenter**, ou tapez **CTRL-MAJ-P** ou appuyez sur le bouton **Indent** (celui-ci : ) si vous ne voyez que le début d'une ligne. Scrollez et regardez bien comment le stylage de ce document bureatique a été pris en charges par la conversion.

- Plusieurs paragraphes utilisent l'attribut *@rend*. Ces valeurs correspondent au stylages utilisés dans le fichier docx ou odt. Donc ils sont disponibles pour effectuer une conversion plus intelligente.

### 3.3 Structure du fichier

Voyons de pres ce que cette conversion aura fait de notre fichier Word. En tout cas, il est devenu un document TEI valide, même s'il pourrait etre amélioré.

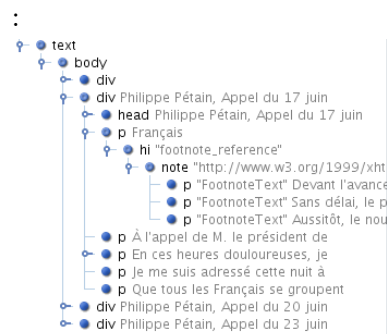
- Relancez Oxygen (si necessaire), et ouvrez le fichier que vous venez de creer.
- D'abord, explorez un peu la structure du fichier. La structuration est plus evidente dans la fenetre Sommaire à gauche de l'ecran. Cliquez à gauche de chaque noeud pour l'ouvrir, ou pour l'enfermer. Par exemple, apres avoir clique à gauche de "text" et de "body", vous allez voir ceci:



- Ensuite, si vous cliquez à gauche du *deuxieme* "div", vous allez voir ceci:



- Vous allez vite comprendre que dans ce fichier, il y a quatre `<div>`, dont le premier ne contient que le titre du fichier. Apres cela, il y a trois `<div>`, contenant un `<head>` et une serie de `<p>`. Quelques-uns de ces `<p>` ne contiennent que de texte, mais il y en a aussi qui contiennent un `<hi>`, contenant un `<note>`, contenant des `<p>` emboites comme ceci :



- Pourquoi? Notre fichier, comme tout autre fichier issu d'un outil bureautique, retient les annotations dans le texte à la point de référence. L'opération de stylage classique sera d'extraire les annotations, les numerotées, et les affichées en bas de page : une espece de transformation.

#### 3.4 Correction du fichier

Avant de continuer, vous etes invité d'améliorer un peu le balisage du fichier. Essayez de :

- transférer les deux titres dans le bon endroit du **<teiHeader>** ;
- enlever la premiere division du fichier ;
- ajouter une numérotation à chaque division du fichier (basé sur la date, p.e. **<div n="1940-06-17" .>**

Une version corrigée est à votre disposition, si vous en avez besoin, dans le fichier **petain-sample-corr.xml**!

#### 3.5 Transformation du fichier

Vous avez deja vu Oxygen faire une transformation de XML : si vous affichez votre document en mode Auteur, vous allez voir que les titres sont maintenant plus jolis, numerotes et mises en evidence, et que les annotations sont affichees avec une police plus petite, mais peut etre pas à l'endroit que vous les attendiez.

Essayons d'autre transformations. Il serait impressionant de transformer notre fichier directement en HTML ou PDF pour l'afficher, n'est-ce pas?

- Sur le barre aux outils, il y a un bouton qui contient une fleche rouge. Cliquez la dessus (ou taper CTRL-MAJ-T, ou selectionnez Document -> Transformation -> Appliquer le scenario de transformation )
- En Oxygen un 'scenario de transformation' associe un document avec une feuille de style, et la fait tourner. Par default, vous voyez **TEI P5 XHTML** et **TEI P5 PDF**. Sélectionnez le scenario qui vous plait, et regardez ce qu'il donne, en cliquant le bouton **Transformer maintenant**. Experimentez-l' autre !
- Bien sur, on peut configurer ces transformations à volonte, et en creer des nouveaux. Pour vous en donner le goût, on va finir en vous proposant une transformation qui n'a rien à faire avec la visualisation.

La feuille de style qui controle une transformation s'ecrit dans un autre langage XML, qui s'appelle XSLT : sujet fascinant, mais peut etre pas pour les debutants. Vous trouverez quelques exemplaires de feuilles de styles XSLT dans votre dossier. Pour en servir, il faut créer une nouvelle transformation en Oxygen

- Cliquez sur la flèche rouge (CTRL-MAJ-T) de nouveau
- Cliquez le bouton **Nouveau** et sélectionnez **XML Transformation with XSLT** du menu déroulant; la fenetre **Nouveau Scenario** s'ouvre. Dans cette fenetre :
  - fournissez un nom pour le scénario : nous proposons "texte-brut"
  - spécifiez le fichier XSL souhaité :
    - \* cliquez sur le petit icone dossier jaune à droite du champs **XSL URL**
    - \* sélectionnez **Parcourir les fichiers locaux** du menu déroulant; la fenetre **Ouvrir** s'ouvre
    - \* sélectionnez le fichier **texte-brut.xsl** et cliquez **Ouvrir** pour fermer ce dialog.
  - sélectionnez **Saxon HE 9** du menu **Transformateur**
  - cliquez **Transformer Maintenant** pour voir le resultat.

Comme vous voyez, c'est possible de transformer notre document XML ou bien dans un format d'affichage, ou bien dans un format d'analyse tres classique.