

Personnalisation de schéma avec Roma

TEI @ Oxford

janvier 2015

1 Objectifs de cet exercice

- Choisir entre les modules TEI
- Combiner les modules TEI pour créer un schéma personnalisé
- Sélectionner un sous-ensemble d'éléments d'un module TEI
- Associer un schéma personnalisé avec un document TEI-XML dans oXygen
- Modifier les valeurs légales des attributs disponibles dans un schéma
- Générer automatiquement la documentation d'un schéma avec Roma
- Se familiariser avec le format TEI ODD XML sous-jacent

2 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 Modèles du Framework, puis sélectionnez TEI P5, enfin cliquez sur All, pour sélectionner un schéma TEI maximal.
- Cliquez sur le bouton Créer en bas de la fenêtre. oXygen affiche alors un nouveau document TEI dans la fenêtre de l'éditeur.
- Enregistrez ce fichier dans le répertoire Travaux sous le nom myTest.xml.

3 All n'est pas une solution idéale

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

- Dans le document que vous venez de créer, placez le curseur à l'intérieur de l'élément `<p>`.
- Tapez `<`. oXygen vous propose dans un menu contextuel tous les éléments disponibles à cet endroit.
- Passez en revue la liste des éléments disponibles : 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 ; puis supprimez le '`<`' que vous venez d'ajouter.

On a beaucoup de choix... mais pour un projet donné, il est peu 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.

4 Création d'un schéma avec Roma

Le logiciel Roma vous permet de créer votre propre schéma TEI. Notre objectif sera de créer un schéma simple pour traiter des sources manuscrites, 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 ici 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 donc **Create** et sélectionnez **TEI Bare**, puis cliquez sur le bouton rouge **start**.
- Pour régler les paramètres on vous propose alors plusieurs choix :
 - *Title* : Changez « TEI absolutely bare » par (par exemple) « TEI pour les manuscrits ».
 - *Filename* : Changez « `tei_bare` » par (par exemple) `teiTranscr` (c'est un identifiant, et donc il ne devrait pas inclure d'espace)
 - *Language* : Vous pouvez laisser l'anglais, mais ce tutoriel part du principe que vous préférez travailler en français. Sélectionnez alors la case à cocher « français ».
 - *Author name* : Entrez votre propre nom (ou, si vous préférez, votre pseudonyme)
 - Modifiez la description si vous le souhaitez.
- Cliquez sur le bouton rouge **Save** en bas de page. Vos changements ont alors 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 affichée.
- Roma est un logiciel international (plus ou moins). Sélectionnez l'onglet **Langage**, et sélectionnez la case à cocher **Français** pour vous assurer que la francisation de cet exercice soit maximale !
- De nouveau, cliquez sur **Save** en bas de page.

5 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 un schéma très minimaliste. Donc, la liste des modules sélectionnés, qui s'affiche à droite, ne contient que trois modules.

- Nous souhaitons disposer des balises pour traiter la transcription des manuscrits. Nous aurons donc besoin d'ajouter au moins deux modules supplémentaires.
- Cliquez sur le mot **ajouter** devant la mention du module **transcr** dans la liste à gauche. Ce module apparaîtra alors dans la liste de droite.
- Répétez cette manœuvre pour ajouter le module **figures**.
- Vous disposez maintenant d'un schéma construit à partir des modules : **tei** (Roma ne vous permet pas de supprimer ce module infrastructurel), **core**, **header**, **figures**, **transcr** et **textstructure**.

6 Inclusion et exclusion des éléments

En dépit de ce faible nombre de modules, il reste malgré tout plusieurs éléments sans 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 alors.
- 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 qui fournit un lien hypertextuel 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. Sélectionnez la case à cocher Exclure dans le titre de la table, et tous les éléments sont exclus par défaut. Il en est de même pour Inclure tous les éléments. Cliquez le bouton à côté d'un élément pour changer son statut !
- Pour cet exercice vous êtes invité à sélectionner que les éléments suivants :
 - Du module **core** vous devez exclure tout les éléments sauf : **<add>**, **<author>**, **<choice>**, **<corr>**, ****, **<graphic>**, **<head>**, **<hi>**, **<lb>**, **<p>**, **<pb>**, **<reg>**, **<sic>**, **<signed>**, **<title>** **<trailer>** et **<unclear>**.
 - Du module **header** vous devez exclure tous les éléments sauf : **<fileDesc>**, **<publicationStmt>**, **<sourceDesc>**, **<teiHeader>** et **<titleStmt>**.
 - Du module **textstructure** vous devez exclure tous les éléments sauf : **<TEI>**, **<body>**, **<div>**, **<salute>**, **<signed>**, **<text>** et **<closer>**
 - Du module **figures** vous devez exclure tous les éléments sauf : **<figure>** et **<figDesc>**.
 - Enfin, du module **transcr** vous devez inclure seulement : **<facsimile>**, **<sur-face>**, et **<zone>**.
- N'oubliez pas de cliquer sur le bouton rouge **Save** en bas de page pour enregistrer vos modifications !

7 Génération d'un schéma

- Cliquez maintenant sur l'onglet **Schéma**. Vous pouvez choisir entre plusieurs langages de schéma : la TEI se définit indépendamment de ces langages, 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 devrait télécharger un fichier de schéma. Placez ce fichier dans votre dossier **Travaux**.
- **Attention ! Ne fermez pas votre navigateur, vous en aurez encore besoin.**

8 Associer un schéma et un fichier XML avec oXygen

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

- Revenez dans oXygen et ouvrez votre fichier
- Dans le menu **Document** ouvrir le menu **Schéma** et sélectionnez **Associer un schéma**. (Remarquez 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 en forme de dossier à côté de **URL** pour naviguer vers le fichier RNG que vous venez de créer avec Roma, puis cliquez sur **OK**.
- Dans votre fichier XML-TEI, vous devriez maintenant voir apparaître une ligne ressemblant à ceci :

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

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

9 Confirmation

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

- Placez de nouveau le curseur après une balise ouvrante `<p>`, et tapez '`<`' pour voir s'ils figurent toujours dans le menu contextuel des éléments disponibles à cet endroit.
- Tous les éléments que vous avez supprimés ont-ils bien disparu ? Alors c'est parfait ! Sinon, peut-être que vous avez oublié de cliquer sur le bouton **Save** après les avoir inclus/exclus... essayez de nouveau !

10 Comment limiter les valeurs de l'attribut *@type* de l'élément `<div>`

Nous avons vu qu'il peut être 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 de façon très permissive : vous pouvez taper n'importe quoi). Supposons que nous désirons limiter les valeurs possibles pour l'attribut *@type* sur `<div>`.

-
- Revenez dans Roma
 - Cliquez sur l'onglet **Modules**
 - Cliquez sur le module **textstructure**
 - Trouvez la ligne où est définie l'élément **<div>** et cliquez sur **Changer les Attributs** tout à fait à droite.
 - Une table des attributs disponibles pour cet élément s'affiche alors. Vous pouvez inclure ou exclure les attributs qui vous intéressent. Pour le moment essayons quelque chose d'autre.
 - Descendez jusqu'à *@type* et cliquez sur l'attribut. Une page s'affiche maintenant, 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 quel 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*) **prose,vers,drama,chapitre,volume,autre,inconnu**.
 - *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 ».
 - **De nouveau, n'oubliez pas de cliquer sur le bouton **Save** en bas de la page !**

11 Testons à nouveau notre schéma...

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

- Cliquez sur l'onglet **Schéma**
- Choisissez l'un des formats de schéma
- Cliquez sur **Générer**
- Trouvez le fichier généré qui a été téléchargé ; copiez-le à la place de l'ancienne version pour le remplacer
- **Surtout, ne fermez pas votre Navigateur Internet !**
- Revenez dans votre fichier **myTest.xml** dans **oXygen** ; choisissez **Document -> Valider -> Validate**

- Si tout va bien, votre document reste valide : il y a un petit carré vert.
- Maintenant il faut ajouter une **<div>**. Comme vous avez fait précédemment, sélectionnez le premier **<p>** dans ce document, et entourez-le d'une balise **<div>** en tapant CTRL-E.
- Votre document n'est plus valide : un carré rouge s'affiche en haut à droite ! En bas vous voyez le message suivant : **Element 'div' missing required attribute 'type'**
- Placez maintenant le curseur à l'intérieur de la balise ouvrante de la **<div>**, juste avant le **>** et tapez un espace. oXygen propose alors une liste des attributs disponibles. Parcourez-la jusqu'à *@type* et notez que cet attribut s'affiche en gras pour signaler que sa présence est requise.
- Sélectionnez *@type* et notez que oXygen vous propose la liste des valeurs que nous avons prédéfinies dans notre schéma.
- Choisissez parmi les valeurs paroposées, par exemple **prose**, et votre document redevient valide, en affichant un petit carré vert.

12 Génération d'une documentation

Tout projet nécessite une 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.
- Remarquez 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.

13 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 cette fois un fichier XML-TEI, du même nom que votre schéma, c'est-à-dire **teiTranscr.xml**. Enregistrez ce fichier dans votre dossier **Travaux**.
- Ouvrez maintenant ce fichier XML avec oXygen pour en prendre connaissance (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 cette option !

14 Pour aller plus loin...

Tester votre schéma d'abord avec un petit manuscrit historique. Ouvrez le fichier `degaulle.xml` qui se trouve dans le dossier travaux et associez ce fichier avec votre schéma. S'il y a des éléments qui manquent dans votre schéma, sauriez-vous comment les ajouter ?

Bien que valide, cette transcription du document original n'est pas très honnête. Les paragraphes de l'original ne sont pas indiqués, et la transcription confond des phrases raturées et des ajouts avec le reste du texte. En vous repérant avec les images `degaulle_recto.jpg` et `degaulle_verso.jpg` contenues dans le même répertoire, améliorez le balisage. Balisez les mots raturés (par exemple 'Ses gouvernantes premieres') avec une ``. De même, balisez les mots ajoutés (par exemple 'cédant à la panique') avec `<add>`. Et bien sûr, indiquez chacun des paragraphes avec l'élément `<p>`.

Ensuite vous pouvez continuer la transcription (nous n'avons fait que la partie initiale).