

TP-3 Personnalisation de schéma avec Roma

TEI @ Oxford

Novembre 2012

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
- Choisissez Cadre des modèles, ensuite TEI P5, puis TEI-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.

3 ALL n'est pas une 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.

4 Création d'un schéma avec Roma

Le logiciel Roma vous permet de créer votre propre schéma TEI.

- 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. **Template** : on part des personnalisations déjà proposées par la TEI
 4. **Open** : on retravaille une personnalisation qu'on a déjà enregistrée.
- Dans cet exercice, nous choisissons la deuxième solution. Choisissez 'reduce' et cliquez sur le bouton rouge 'start'.
- Pour régler les paramètres on vous propose :
 - **Title** : Changez 'TEI with maximal setup' en (par exemple) "TEI pour les manuscrits modernes".
 - **Filename** : Change 'tei_all' en (par exemple) "monTEI" (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)
 - Ne changez pas la description pour l'instant.
- Cliquez sur le bouton **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.

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 le schéma maximal. Donc, la liste des modules sélectionnés, affichée à droite, correspond exactement à la liste des modules disponibles, affichée à gauche.
- Cliquez sur 'remove' près de 'analysis' sur la liste affichée à droite. Le module disparaît de cette liste, mais reste toujours disponible dans la liste à gauche. Vous pourriez le remettre si besoin est.

- Enlevez 'analysis', 'certainty', 'corpus', 'dictionaries', 'drama', 'figures', 'gaiji', 'iso-fs', 'linking', 'nets', 'spoken', 'textcrit', 'verse', et 'tagdocs' !
- Vous avez maintenant un schéma construit à partir des modules : 'tei' (Roma ne vous permet pas d'enlever ce module infrastructurel), 'core', 'header', 'msdescription', 'namesdates', 'textstructure', et 'transcr'. Ces modules nous semblent les plus pertinents pour travailler avec les manuscrits modernes.

6 Inclusion et exclusion des éléments

Dans ce petit ensemble de modules, il reste quand même plusieurs éléments sans intérêt pour notre projet.

- Cliquez sur le mot 'core' à droite (nota : non pas sur le mot 'remove' 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
- Vous pouvez Inclure ou Exclure tous les éléments dans un module. Cliquez sur le mot Exclure dans le titre de la table, et tous les éléments sont exclus par défaut. Puis cliquez de nouveau Inclure.
- Du module 'core' on va exclure les éléments suivants : 'addrLine', 'address', 'analytic', 'biblStruct', 'binaryObject', 'distinct', 'divGen', 'gb', 'headItem', 'headLabel', 'imprint', 'index', 'listBibl', 'measure', 'measureGrp', 'meeting', 'mentioned', 'monogr', 'postBox', 'postCode', 'relatedItem', 'rs', 'said', 'series', 'sp', 'speaker', 'stage', 'street', 'teiCorpus', 'term', 'textLang', and 'time'.
- Notre schéma devient un peu plus gérable.... **N'oubliez pas de cliquer sur le bouton rouge 'Save' en bas de page !**
- Dans un projet réel, bien sûr, on continuerait de la même manière pour chacun des modules, afin de définir un sous-ensemble d'éléments correspondant exactement aux besoins de notre projet. Plus le schéma est restreint, plus nos données seront consistantes !

7 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. Vous devez enregistrer ce fichier dans le même dossier que le fichier myTest.xml que vous avez créé au début de cet exercice.
- Ne fermez pas votre navigateur : vous en aurez encore besoin.

8 Association de schéma et fichier en oXygen

Jusqu'à présent, oXygen associait le schéma `tei_all` avec votre fichier `myTest.xml` (c'est le schéma sélectionné à cause de la présence d'un élément TEI dans l'espace de noms TEI).

- Revenez dans oXygen et ouvrez 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 href="myTEI.rng"
type="application/xml"
ns="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 ?

- 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 !

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

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 `<div>`.

- Revenez dans Roma
- Cliquez sur l'onglet 'Modules'
- Cliquez sur le module 'textstructure'
- Trouvez le rang où est défini l'élément 'div' 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**) **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".

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

11 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 (cache)'
- 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 avant, sélectionnez le premier **<p>** dans ce document, et entourez-le d'une balise **<div>** 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 'div' 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. 'prose', et votre document devient valide, avec un petit carré vert.

12 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 'Generate'.
- 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.

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 un fichier XML-TEI, avec le même nom que votre schéma, donc `myTEI.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 TEI All 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. 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 !

14 Limitations de Roma

En effet, Roma n'est qu'une interface web basée sur un système de transformations TEI beaucoup plus complexes, que vous avez déjà rencontrées avec oxGarage. Il est possible de créer son fichier de personnalisation sans se servir de Roma : c'est une démarche qui permet, par ailleurs, de faire beaucoup plus de choses qui sont rendues impossibles par l'interface de

Roma. Par exemple, Roma vous permet de limiter les valeurs légales d'un attribut, comme vous l'avez vu, mais ne vous permet pas de définir et de documenter la portée de ces valeurs ; chose qui est simple à faire en manipulant directement le format XML-TEI de votre personnalisation (votre 'ODD').

15 Pour en savoir plus

Voici quelques lectures intéressantes à ce sujet :

Documentation de référence (dans les Guidelines)

- Description complète du langage ODD : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>.
- Description du traitement des fichiers ODD : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#IM>

Autres documents tutoriels

- <http://tbe.kantl.be/TBE/modules/TBED08v00.htm>
- <http://www.tei-c.org/Guidelines/Customization/odds.xml>