Technique et chaîne de publication électronique avec XSLT

Jean-Damien Généro 2023, 5 dec. - 2024, 23 janv.

École nationale des chartes - M2 TNAH

Contact

· jean-damien.genero@cnrs.fr

Ί

Objectifs

XPath

- Naviguer dans un arbre XML
- Manipuler les principales fonctions XPath

XSLT

- · Manipuler les règles (templates) basiques
- · Manipuler les conditions et les boucles

Édition numérique

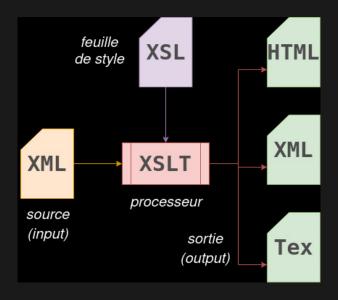
- Transformer un document XML en un autre document XML
- Transformer un document XML en un document HTML
- Transformer un document XML en un document धा_EX

Introduction. Écosystème XML

Écosystème XML/ Principes généraux

- XML → langage de balisage qui encode une description de la mise en page et de la structure logique d'un document
 - XPath → langage de requête pour naviguer dans la structure hiérarchique d'un document XML à l'aide de chemins;
- XSL → spécifications pour écrire des feuilles de style (stylesheet);
- XSLT → un langage conçu pour transformer des documents XML en d'autres documents selon les spécifications XSL.

Écosystème XML/ Schéma



Écosystème XML/ Navigation XPath et de règles XSL

Navigation XPath :

- Doc XML = structuré, on parle d'« arbre » (XML tree);
- Itinéraire vers une balise (« parent » → « enfant »);
- · Rédaction dans une syntaxe propre.

· Règle XSL (template) :

- Agir sur une ou plusieurs balises (transform);
- Exemples: copier balise + contenu, copier uniquement le contenu et l'insérer dans une nouvelle balise, ajouter ou supprimer un attribut, etc.;
- · Rédaction dans une **syntaxe XML** (langage à balises).

Utilisation avec d'autres langages (Python: lxml).

Écosystème XML/ Exemple d'un chemin XPath dans un XML tree

Comment accéder aux balises avec XPath?

```
title
text
```

- · <TEI> \rightarrow <text> \rightarrow <body> \rightarrow <div> \rightarrow
- Traduction XPath : /TEI/text/body/div/p
- Simplification: //body/div/p (/TEI/text/body/div/p)

Écosystème XML/ Exemple d'une règle XSL

- Comment appliquer une règle XSL aux balises ?
 - Une règle XSL est toujours contenue dans une balise <xsl:template/> possédant un @match.
- · Indiquer le chemin XPath vers les dans l'*∂match*.

ou

```
<xsl:template match="//body/p">
     <xsl:copy-of select="."/>
</xsl:template>
```

1. XPath

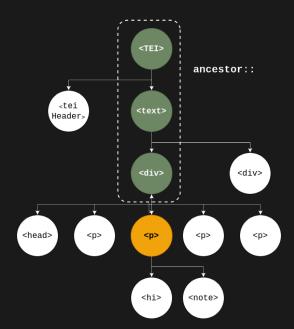
XPath/ Nœuds

- Nœud (node) (7) = composant de l'arbre : root, element, attribute, text, comment, namespace et processing instruction.
- Deux rôles : current node (fixe, le premier du chemin) ou context node (variable, celui que XPath évalue à l'instant t);
- Écrire un chemin (expression): succession de nœuds séparés par l'opérateur /;
 - Tester dans Oxygen : /TEI/text/body/div/p

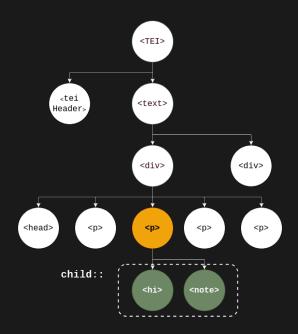
XPath/ Axes de relation

- Un axe permet de qualifier la relation entre le nœud de contexte et un ou des autre(s) nœud(s).
- Direction basique : parent → enfant.
- 13 axes XPath, dont :
 - parent:: → nœud immédiatement au-dessus;
 - child:: → nœud immédiatement en-dessous;
 - following-sibling:: → nœud(s) après celui de contexte et qui ne fait/font pas partie de ses descendants;
 - ancestor, descendant, preceding, etc.

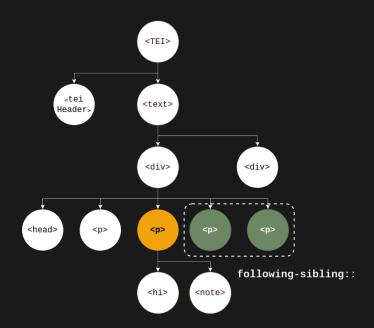
XPath/ Axe ancestor::



XPath/ Axe child::



XPath/ Axe following-sibling::



XPath/ Abbréviations des axes

- 1er élément de l'arbre (root node) → /
 (attention : différent de la racine <TEI>)
- decendant-or-self:: ightarrow //
- $self:: \rightarrow$.
- $oldsymbol{\cdot}$ attribute::ightarrow @
- Donc : vous pouvez écrire //div/@n au lieu
 de //div/attribute::n.

XPath/ Prédicats: définition

- Prédicat (filtre) = condition qui doit être satisfaite par le nœud de contexte (existence d'un attribut, valeur d'un attribut, position d'une balise, etc.).
- Écrit entre crochets [].
- · //div[@n='2']/head = le <head> de la <div> avec un @n de valeur 2.
- ·//body/div[@n='2']/p[2]=?

XPath/ Prédicats : exemples

- //tag[position()=2] ou //tag[2] →
 condition de position (<tag> n°2);
- //tag[last()] → dernier <tag>;
- //tag[@foo='bar'] → expression logique (<tag> avec un @foo dont la valeur est bar);

XPath/ Notions essentielles

- Arbre XML, racine, nœuds (7, notés avec ());
- Chemin XPath, opérateur /;
- Axes de relation (13, notés avec ::);
- · Prédicats (notés avec []).

instructions de premier niveau

2. XSLT : élément racine et

XSLT/ Définition

- XSLT → un langage de programmation;
- Permet de transformer un doc XML en un autre doc (.xml, .html, .tex, etc.);
- Transformation opérée par un processeur XSLT
 - Construit l'arbre output, transforme l'arbre input et sérialise le document output;
 - Le plus connu → Saxon (en ligne de commande);
 - Généralement intégré à des logiciels (Oxygen) ou des librairies (python: lxml).

XSLT/ La feuille de style

- Feuille de style XSL == un doc XML .xsl;
- Contient des instructions ou règles (templates);
- Élément racine : <xsl:stylesheet>

- Dans Oxygen \rightarrow ouvrir un nouveau doc
 « XSLT Stylesheet ».
- · Observer les attributs de l'élément racine.

XSLT/ Élément racine : <xsl:stylesheet> (1/4)

Espace de noms et attributs présents de base dans Oxygen :

- @xmlns:xsl: espace de noms (namespace) XSL;
- อxmlns:xs: espace de noms XML;
- Dexclude-result-prefixes: liste des préfixes qui ne seront pas copiés dans l'output;

XSLT/ Élément racine : <xsl:stylesheet> (2/4)

Attributs à remplacer dans Oxygen pour une transformation vers XML-TEI :

- remplacer ∂xmlns:xs par ∂xmlns:tei → espace de noms de l'élément racine de l'output.
 - évite l'ajout de *tei*: à chaque élément "matché" par une règle (*@match*).
- remplacer xs par tei dans ∂exclude-result-prefixes.

XSLT/ Élément racine : <xsl:stylesheet> (3/4)

Attributs à ajouter dans Oxygen pour une transformation vers XML-TEI :

- axpath-default-namespace: espace de noms des chemins XPath de la feuille de style (http://www.tei-c.org/ns/1.0).
 - évite de devoir écrire le préfixe *tei*: devant chaque nœud d'une expression XPath.
- axmlns avec l'adresse de l'espace de noms TEI.
 - détermine l'espace de noms de l'ensemble du document de sortie.

XSLT/ Élément racine : <xsl:stylesheet> (4/4)

- attention: TEI est un standard XML parmi d'autres! Pour EAD, changer l'espace de noms TEI par celui d'EAD et le préfixe tei: par ead:
- Atributs pour la transformation vers HTML ou LaTeX :
 - อxmlns:xsl (adresse de l'espace de noms XSLT)

 - · *᠗version* (version de XSL utilisée)

XSLT/ Instruction de 1er niveau: <xsl:output>

- · Instruction de premier niveau;
- « contrôle les caractéristiques du document de sortie ».

```
<xsl:output
    method="xml | html | text"
    indent="yes | no"
    omit-xml-declaration="yes | no"
    encoding="UTF-8"
/>
```

XSLT/ Écriture d'une règle : <xsl:template> (1/2)

- une règle peut être écrite de plusieurs manières;
- <xsl:template> → définir une règle;
- Possède un @match avec pour valeur un chemin XPath qui donne l'emplacement du nœud sur lequel sera appliquée la règle.

XSLT/ Écriture d'une règle : <xsl:template> (2/2)

 Effet d'une règle vide → partie de l'arbre sélectionnée n'est pas copiée dans la sortie.

```
<xsl:template match="/">
</xsl:template>
```

 Effet de <xsl:copy-of/> → copie à l'identique la balise matchée par le @select et ses nœuds enfants.

XSLT : première approche. Notions essentielles

- Transformation et feuille de style XSL;
- Manipulation de XSL avec Oxygen;
- Élément racine : <xsl:stylesheet> et ses attributs (en-tête XSL);
- · Instructions de premier niveau :
 - . <xsl:output>;
 - ・<xsl:template> et son @match.

3. XSLT. Règles basiques

XSLT/ <xsl:template>

- Un <xsl:template> contient une ou plusieurs règles;
- L'*amatch* contient le chemin XPath vers la balise qui sera le point de départ de la ou des règles, qui peut s'appliquer à :
 - La balise désignée dans le @match;
 - · Les enfants ou descendants de cette balise;
 - N'importe quelle balise de l'arbre via des axes (ancestor, preceding, etc.).
 - La partie de l'arbre sur laquelle la règle s'applique est en général indiquée dans le aselect de l'instruction XSL (cf. infra).

XSLT/ Contenu d'un <xsl:template>

- · <xsl:template> peut contenir:
 - Des balises XML ou HTML (avec leurs @tt);
 - Du texte (commandes \(\mathbb{T}_EX\);
 - Des instructions ou variables/paramXSL : <xsl:.../>;

Contenu d'un <xsl:template> : fichier d'exemple

- Ouvrir Oxygen.
- Désactiver le paramètre par défaut :
 Preferences > XML > XML Parser > RELAX NG,
 décocher Add default attributes
- Ouvrir class2_teiAll_oxygen.xml.
- Ouvrir une feuille de style XSLT vide, compléter l'en-tête XSLT et ajouter l'instruction <xsl:output/> (cf. fiche xsl_header).

XSLT/ Contenu d'un <xsl:template> : XML, HTML, 上TFX (1/2)

 Des balises peuvent être écrites directement dans un <xsl:template>:

```
<xsl:copy-of select="./p[2]"/>
```

XSLT/ Contenu d'un <xsl:template> : XML, HTML, LETEX (2/2)

 Les balises XML ou HTML peuvent être écrites dans des <xsl:element name="tag"> et les attributs dans des <xsl:attribute name="@tt"> :

XSLT/ Contenu d'un <xsl:template> : du texte (1/2)

- Du texte peut être mis dans une règle : pour remplacer le contenu d'une balise, pour écrire le préambule d'un doc ﷺ.
- (Dans le fichier d'exemple : remplacer tout.)

```
<xsl:template match="/">
    \documentclass[a4paper]{book}
    \usepackage[utf8]{inputenc}
    \usepackage[french]{babel}
    \usepackage{fontspec}
    \begin{document}
    \end{document}
```

XSLT/ Contenu d'un <xsl:template> : du texte (2/2)

• Du texte peut aussi être mis dans un <xsl:text>text</xsl:text>:

```
<xsl:attribute name="type">chap</xsl:at</pre>
<xsl:text>nouveau texte</xsl:text>
```

Règles basiques : notions essentielles

- Les balises XML et HTML peuvent être écrites en clair dans un <xsl:template>;
- Les balises XML et HTML peuvent aussi être écrites dans un <xsl:element name="tag"> et un attribut dans un <xsl:attribute name="@att">.
- Le texte (commandes 哲氏) peut être mis soit en clair soit dans un **<**xsl:text>.

4. Les quatre principales instructions

XSL

XSLT/ Les quatre principales instructions XSL

- . <xsl:copy/>
- . <xsl:copy-of/>
- ·<xsl:value-of/>
- · <xsl:apply-templates/>

XSLT/ Principales instructions (1/4): <xsl:copy>

- Copie de la balise matchée par le @match,
 sans les namespaces, attributs, texte, etc.
- <xsl:copy/> peut contenir d'autres règles, du texte, etc.
- Intérêt → copier un élément pour appliquer des règles à ses enfants.

```
XSLT/ Principales instructions (2/4): <xsl:copy-of>
```

- Copie à l'identique de la balise matchée par le @select et de ses nœuds enfants (balises, attributs, textes).
- Élément vide (pas de règles internes).
- Impossible de modifier les éléments copiés!
- Intérêt → reproduire rapidement une partie de l'arbre que l'on ne veut pas modifier.

```
XSLT/ Principales instructions (3/4): <xsl:value-of>
```

- Renvoie uniquement la valeur textuelle de la balise matchée par le @select.
- · Élément vide (pas de règles internes).
- Intérêt → copier du texte sans les balises (utile pour ﷺ).

XSLT/ Principales instructions (4/4)
<xsl:apply-templates/>

- Ouvrir les fichiers
 class2_st_julien_hospitalier.xml
 et att_apply_templates_class2.xsl.
- <xsl:apply-templates/> est une instruction récursive : le processeur va examiner les nœuds enfants de la balise matchée par le @match dans l'ordre et appliquer les règles qui leur sont associées.

XSLT/ L'élément <xsl:apply-templates/> (1/4)

- Sans elle, le processeur s'arrête à l'emplacement désigné par le @match du <xsl:template/> et ne traite pas les éléments enfants.
- Exemple : la balise <TEI/>.

```
XSLT/ L'élément <xsl:apply-templates/>:@select(2/4)
```

- *@select* permet d'appliquer les règles uniquement au nœud sélectionné.
- Ex. → inverser < teiHeader/> et < text/>
 (cf. fichier):

XSLT/ L'élément <xsl:apply-templates/>:@mode(3/4)

- Omode permet d'appliquer des règles différentes à un même élément XML en fonction de son emplacement ou de son contenu (= son « mode ») sans avoir à mettre ces règles dans un seul <xsl:template/>.
- Le @mode doit être présent et dans le <xsl:apply-templates/> et dans le <xsl:template/> avec la même valeur (= le nom du mode).

XSLT/ L'élément <xsl:apply-templates/>:@mode(4/4)

• Exemple d'utilisation de **amode** : créer une table des matières (*cf.* fichier).

```
<xsl:template match="//body">
          <head>Table des matières
```

XSLT/ Une autre instruction: <xsl:number/> (1/2)

- <xsl:number/> « compte des éléments de façon séquentielle ».
- Attributs :
 - acount → définit les éléments de l'input qui seront numérotés dans l'output;
 - alevel="single|multiple|any" → niveaux de l'arbre pris en compte pour le comptage;
 - $\partial format = "1/01/A/a/I/i" \rightarrow format$ des numéros.

XSLT/ Une autre instruction: <xsl:number/> (2/2)

```
<xsl:template match="//body//p">
                     count="//div[@n]/p"
                     level="any"
                     format = "1"/>
            </xsl:attribute>
            <xsl:value-of select="."/>
        </xsl:copy>
```

XSLT/ Ordre d'application des règles

- XSLT commence par chercher la règle à appliquer au nœud racine;
- Cette règle fait appel à d'autres avec apply-templates (avec ou sans @tt), elles sont appliquées dans l'ordre;
- · S'il n'y a pas de règle, il passe au suivant.
- Donc : l'ordre d'écriture des règles n'a aucune importance;
- Attention → restez lisible + commentez votre code!

5. XPath/ Les fonctions

XPath/ Fonctions

- Les fonctions peuvent être utilisées dans des expressions XPath ou dans des prédicats :
 - Expression → count(./p)
 - Prédicat ([]) \rightarrow ./[count(p)]
- Cf. fiche: fichier xpath-nodes-axis-functions.

6. XSLT. Variables et paramètres

XSLT/ Différence entre variable et param

- Rappel → variable : nom (unique) + valeur (statique ou dynamique).
- Dans XSLT, portée globale (premier niveau)
 ou locale (intégrée à un xsl:template);
- <xsl:variable/>: valeur fixe (NB : la valeur courante peut changer au sein d'une itération);
- <xsl:param/>: valeur fixe ou dynamique,
 peut être pris en argument par le
 processeur (ex: saxon dans lxml).

```
XSLT/ <xsl:variable/>(1/2)
```

- Attributs : @name (req.) et @select (opt.);
- Deux manières de définir la valeur de
 xsl:variable/>:
 - Par la valeur de l'aselect (expression XPath);
 - Par le contenu de <xsl:variable/>
 (texte ou règle).

```
<xsl:variable name="x" select="Xpath"/>
<xsl:variable name="y">[règles]</xsl:variable>
```

XSLT/ <xsl:variable/>(2/2)

• Pour appeler une variable : \$ + @name dans le @select d'un <xsl:value-of/> (uniquement le texte) ou d'un <xsl:copy-of/> (texte + balises).

XSLT/ <xsl:param/>

 Fonctionne de la même manière que <u style="list-style-type: square;">xsl:variable/>. 7. XSLT. Conditions, boucles et tri

XSLT/ Conditions

- Deux moyens d'exprimer une condition en XSLT : <xsl:if/> et <xsl:choose/>;
- Attention : les conditions XSLT ne fonctionnent pas de la même manière que celles d'autres langages de programmation, comme Python où l'on trouve le triplet if, elif et else.

XSLT/ <xsl:if/> (1/2)

- <xsl:if/> possède un atest obligatoire;
- Contenu appliqué si le **atest** est validé.
- Pas de elif: utiliser plusieurs <xsl:if/>;
- Pas de else: utiliser <xsl:choose/>.

XSLT/ < xsl:if/> (2/2)

Exemple : dans St Julien, ajouter un ∂n aux $> si ces derniers sont dans la <math>< div/> n^{\circ} 2$:

```
<xsl:template match="//div[@n]/p">
        <xsl:value-of select="."/>
```

XSLT/ <xsl:choose/>

- <xsl:choose/> a pour enfant un ou des
 <xsl:when/> (req.) et un seul
 <xsl:otherwise/> (opt.).
- · <xsl:when/> a un @test obligatoire
 (même fonctionnement que <xsl:if/>);
- si aucun <xsl:when/> n'est réalisé, alors <xsl:otherwise/> est activé

XSLT/ Boucle simple : <xsl:for-each/> (1/2)

- <xsl:for-each/> a un @select;
- Le point de départ des chemins Xpath au sein d'un <xsl:for-each/> est la balise sélectionnée par le @select;
- Peut intégrer d'autres instructions XSL comme <xsl:if/> et <xsl:choose/>.

```
XSLT/ Boucle simple : <xsl:for-each/>(2/2)
```

• Exemple:

XSLT/ Tri: <xsl:sort/> (1/2)

S'utilise comme premier enfant de
<xsl:for-each/> ou
<xsl:for-each-group/>;

 Change l'ordre d'origine des nœuds sélectionnés en un autre ordre (alphabétique, numérique, etc.);

```
XSLT/ Tri: <xsl:sort/> (2/2)
```

• Exemple :

XSLT/ Itération sur des groupes : <xsl:for-each-group/>
(1/3)

- Rassemble les nœuds (@select) en groupe selon un critère donné (@group-by) et leur applique les règles définies à l'intérieur de <xsl:for-each-group/>.

 - Regroupe ces nœuds selon le critère défini dans le @group-by;
 - 3. Applique les règles indiquées dans le contenu de la balise.

XSLT/ Itération sur des groupes : <xsl:for-each-group/>
(2/3)

- Plusieurs <xsl:for-each-group/> sont souvent utilisés ensemble (boucles dans une boucle);
- Ils peuvent contenir des <xsl:for-each/>, des conditions, etc...
- · Attention : c'est une nouveauté de XSLT 2.0.

XSLT/ Itération sur des groupes : <xsl:for-each-group/>
(3/3)

- current-group() retourne la liste des nœuds du groupe de l'itération en cours d'exécution;
 - à utiliser dans le @select d'un sous-<xsl:for-each-group/> ou d'un <value-of/>.
- current-grouping-key() retourne la clef utilisée pour l'itération en cours d'exécution.
 - à utiliser dans le @select d'un <value-of/>.

XSLT/ Conditions et boucles. Notions essentielles

- Conditions XSL :
 - Sans alternative: <xsl:if/> (@test);
 - 2. Avec alternative : <xsl:choose/>, contient des
 <xsl:when> (@test) et un
 (<xs:otherwise/>).
- Boucles XSL:
 - 1. Simple: <xsl:for-each/>(@select);
 - 2. Sur des groupes : <xsl:for-each-group/> (@select, @group-by).
- ・Tri XSL: <xsl:sort/>(@select).

8. Appel d'une template : <xsl:call-template/>

XSLT/ Instruction call-template (1/2)

- · <xsl:call-template name="#"/>;
- Permet d'appeler autant de fois qu'on le veut une template dans un
 - <xsl:template/>
- · Cas pratiques :
 - Un call-template par result-document;
 - Peut permettre de stocker des parties du document de sortie (header, footer) au même titre que les variables.

XSLT/ Instruction call-template (2/2). Exemple.

```
<xsl:template match="/">
    <xsl:call-template name="name1"/>
</xsl:template>
    <xsl:result-document href="#">
        <html>
        </html>
    </xsl:result-document>
</xsl:template>
```

9. XSLT. Enregistrement du document de sortie

XSLT/ Enregistrement : <xsl:result-document/> (1/2)

- <xsl:result-document/> permet de générer plusieurs documents de sortie;
- Nouveauté de XSLT 2.0;
- S'utilise comme enfant d'un <xsl:template/> ou d'un <xsl:call-template/>.

XSLT/ Enregistrement: <xsl:result-document/>(2/2)

Trois attributs:

- ahref permet d'indiquer l'URI du document à générer : absolue (depuis la racine) ou relative (dans ce cas, le processeur part de l'emplacement du fichier d'entrée);
- amethod permet de spécifier le format du code de sortie;
- *aindent* permet d'indiquer si le résultat sera indenté;

XSLT/ <xsl:result-document/>, exemple

```
<xsl:template match="/">
        method="xml" indent="yes">
    <TEI><!-- règles --></TEI>
        method="xml" indent="ves">
```

XSLT/ <xsl:result-document/> et variables

- Inconvénient de créer X docs de sortie → répéter X fois des éléments récurrents (<teiHeader/>, header HTML, etc.);
- Stocker ces éléments dans des variables permet de ne les modifier qu'une seule fois.

XSLT/ <xsl:result-document/> et URI

 Bonne pratique de programmation → déclarer les URI des docs de sortie dans des variables au début de la feuille de style.

```
<xsl:template match="/">
```

XSLT/ <xsl:result-document/>. Notions essentielles

- <xsl:result-document/> permet de générer un ou plusieurs documents de sortie;
- L'URI est indiquée dans un **@href** ;
- Les variables sont utiles pour :
 - Déclarer les URI des documents de sortie au début de la feuille de style;
 - Contenir les éléments récurrents des documents de sortie (header, footer, etc).

10. XSLT. Transformation vers HTML

XSLT/ Méthode HTML (1/2)

- Dans l'en-tête XSL : les @xmlns:tei,
 @xmlns et @exclude-result-prefixes
 ne sont plus nécessaires;
- Dans <xsl:method/>: ajouter
 @method='html',
 @omit-xml-declaration n'est plus nécessaire.

XSLT/ Méthode HTML (2/2). Template HTML basique

```
<html lang="fr">
    <head>
        <title><!-- doc title --></title>
        <meta name="description" content="#"/>
        <meta name="author" content="#"/>
        <!-- <link rel="stylesheet" href="#"/> -->
    </head>
    <body>
        <!-- some text here -->
        <!-- <script src="#"></script> -->
    </body>
                                                  73
</html>
```

11. XSLT. Transformation vers 上下X

XSLT/ Méthode texte (1/2)

- Même en-tête XSL que pour la transformation vers HTML.
- Transformation directe de XML à শEX
 non-prise en charge par XSLT → utiliser la
 méthode text dans le <xsl:output/>.
- Conséquence → les commandes धा_EX ne sont pas reconnues comme telles et traitées comme du texte.

XSLT/ Méthode texte (2/2). Template 上X basique

```
\documentclass[]{book}
\usepackage[T1]{fontenc
\usepackage[french]{babel}
\begin{document}
\addcontentsline{toc}{chapter}{Table des matières}
\end{document}
```