

Interoperabilité des projets TEI: chimère ou apothéose?

Lou Burnard

La TEI facilite un balisage 'intelligent'

La TEI s'applique à l'encodage des...

- composants structuraux et fonctionnels d'un texte
- transcriptions diplomatiques des sources historiques, des images, des annotations
- liens, correspondances, alignements
- données et entités : par exemple de temps, personnes, lieux ou événements
- annotations peritextuelles et métatextuelles (correction, suppression, ajouts)
- analyses linguistiques
- métadonnées de plusieurs types
- ... et définitions formelles de schéma XML !

On est vraiment gâté de choix

Les TEI Guidelines

Dans ses 1 400 pp imprimées, vous trouverez :

un lexique 521 éléments, regroupés en 146 classes, et 22 modules

des règles d'usage 7 185 lignes de règles syntaxiques en RELAXNG

des contraintes formalisées 21 types de données, plusieurs règles formalisées en Schematron

des règles / conventions d'utilisation beaucoup de prose

plusieurs exemples d'utilisation dont au moins un par élément

comment régler ses richesses?

Le cornucopia de balises TEI



Liste des modules TEI

analysis	Simple analytic mechanisms
certainty	Certainty and uncertainty
core	éléments common to all TEI documents
corpus	Header extensions for corpus texts
declarefs	Feature system déclarations
dictionaries	Printed dictionaries
drama	Performance texts
figures	Tables, formulae, and figures
gaiji	Character and glyph documentation
header	The TEI Header
iso-fs	Feature structures
linking	Linking, segmentation and alignment
msdescription	Manuscript Description
namesdates	Names and dates
nets	Graphs, networks and trees
spoken	Transcribed Speech
tagdocs	Documentation of TEI modules
tei	déclarations for datatypes, classes, and macros available to all TEI modules
textcrit	Text criticism
textstructure	Default text structure
transcr	Transcription of primary sources
verse	Verse structures

Comment effectuer notre choix de balises?

Nous aurons besoins de plusieurs choses :

- un schéma formel (en langue informatique tel que DTD, RELAXNG, W3C Schema, Schematron) qui peut contrôler :
 - quelles balises sont disponibles ?
 - dans quels contextes ?
 - avec quels attributs ?
 - avec quelles valeurs ?
 - en respectant quelles contraintes ?
- une documentation pour expliquer nos principes éditoriaux, nos principes de choix de balises, etc. aux utilisateurs/ développeurs :
 - en plusieurs langues naturelles
 - en plusieurs formats bureautiques (PDF, Word, HTML, epub...)
- des outils informatiques pour transformer et gérer nos données XML

Propositions de la TEI

Ayant elle-même ces mêmes besoins, la TEI vous propose :

ODD Un vocabulaire XML pour définir les vocabulaires XML

Roma Un logiciel pour traiter des documents exprimés dans ce vocabulaire, pour en faire ressortir

- des schémas RELAXNG, DTD, etc.
- des manuels "mode d'emploi" en HTML, PDF etc.

OxGarage Un outil web pour effectuer plusieurs transformations documentaires, par ex

- de DOCX en TEI, et l'inverse
- de HTML en TEI, et l'inverse

Stylesheets Un ensemble de feuilles de style XSLT très générique pour la conversion des documents XML TEI

I18n

- Suite à un effort financé par le ALLC (entre autres), les définitions, les gloses, et quelques exemples sont traduits en plusieurs langues, notamment le français, mais aussi:
 - espagnol, portugais, allemand, chinois, koreen, japonais...
- Les chapitres des Guidelines actuels restent jusqu'à présent en anglais.
- Mais on a mis en place une architecture pour faciliter sa traduction...

L'idée essentielle de ODD

One Document Does it all

Un vocabulaire spécialisé pour la définition

- des schémas
- des types d'élément XML, indépendant des schémas
- des regroupements de tels éléments, publics ou privés
- des patrons (MLE macros)
- des classes (et sous-classes) d'éléments

Pour la définition des références également, qui permettront de réunir dans un schéma

- des objets identifiables (dans la liste ci-dessus)
- des objets appartenant à d'autres espaces de nom

et qui serait intégrable à un système de balisage documentaire classique

Traitement des ODD

Un processeur ODD :

- rassemble les composants référencés ou fournis
- résoud les incohérences ou duplications éventuelles
- peut contrôler ou valider les résultats
- émet un schéma exprimé en langue formelle (RELAXNG, DTD, WSD...)
- émet un document XML "pur" avec les composants documentaires

<http://www.tei-c.org/Roma>

Choisir sélectivement (1)

Vous pouvez spécifier les éléments que vous souhaitez supprimer parmi ceux qui sont proposés par un module:

```
<schemaSpec start="TEI" ident="testschéma-4a">  
  <moduleRef key="core" except="mentioned quote said"/>  
  <moduleRef key="header"/>  
  <moduleRef key="textstructure"/>  
  <moduleRef key="dictionaries" except="oRef"/>  
</schemaSpec>
```

ou également :

```
<schemaSpec start="TEI" ident="testschéma-4b">  
  <moduleRef key="core"/>  
  <moduleRef key="header"/>  
  <moduleRef key="textstructure"/>  
  <elementSpec ident="mentioned" mode="delete"/>  
  <elementSpec ident="quote" mode="delete"/>  
  <elementSpec ident="said" mode="delete"/>  
</schemaSpec>
```

Choisir sélectivement (2)

Vous pouvez spécifier les éléments que vous souhaitez inclure parmi ceux qui sont proposés par un module:

```
<schemaSpec start="TEI" ident="testschéma-4b">  
  <moduleRef key="core"/>  
  <moduleRef key="header"/>  
  <moduleRef key="textstructure" include="body div"/>  
</schemaSpec>
```

ou également :

```
<schemaSpec start="TEI" ident="testschéma-4b">  
  <moduleRef key="core"/>  
  <moduleRef key="header"/>  
  <elementRef key="div"/>  
  <elementRef key="body"/>  
</schemaSpec>
```

Spécification des listes de valeurs

Les valeurs légales d'un attribut peuvent être spécifiées par un `<datatype>` et/ou par `<valList>`.

Un besoin assez commun est de spécifier une énumération (une liste -- ouverte ou fermée -- des valeurs)

```
<attDef ident="status">
  <desc>indique l'état courant du système selon un code de couleur</desc>
  <defaultVal>green</defaultVal>
  <valList type="closed">
    <valItem ident="red">
      <desc>fermeture complète du système</desc>
    </valItem>
    <valItem ident="orange">
      <desc>fermeture imminente du système</desc>
    </valItem>
    <valItem ident="green">
      <desc>état normal du système</desc>
    </valItem>
    <valItem ident="white">
      <desc>état inconnu du système</desc>
    </valItem>
  </valList>
</attDef>
```

Les classes TEI

Pour modifier un schéma, il faut comprendre la système de classes TEI.

- une classe d'attribut comme att.xxxx est un fournisseur d'attributs
- une classe de modélisation comme model.xxx regroupe des éléments selon leur sémantique, ou leur possibilités d'emplacement :
 - model.xxxLike : pareille à xxx
 - model.xxxPart : constituant un xxx

par exemple

Google [PDR] Athanasius Kircher Text Encoding Initiative TEI class modelGramPart

www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-modelGramPart.html

Google Maps The Usual Press This Facebook bit.ly Sidebar Zoho Invoice BNC Simple Sea... OxLIP+ - Find D... BerneCourse Google Maps Travel

TEI <Text Encoding Initiative>

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — Français

P5: Recommandations pour l'encodage et l'échange de textes

Version 2.3.0. Last up

model.gramPart

Accueil
A Class

model.gramPart regroupe des éléments autorisés dans l'élément [gramGrp](#). [\[9.3.2 Grammatical Information\]](#)

Module	dictionaries — 9 Dictionaries
Utilisé par	gramGrp model.formPart
Membres	model.morphLike [gram gen number case per tns mood iType] gramGrp pos subc colloc usg lbl

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)

TEI Consortium | Commentaires

TEI material can be licensed differently depending on the use you intend to make of it. Hence it is made available under both the CC-BY and BSD-2 licences. The CC-BY licence is generally appropriate for usages which treat TEI content as data or document, generally appropriate for usage of TEI content in a software environment. For further information or clarification, please contact the [TEI Consortium](#).

1. <http://creativecommons.org/licenses/by/3.0/>
2. <http://www.opensource.org/licenses/BSD-2-Clause>

TEI Guidelines [Version 2.3.0](#). Last updated on 17th January 2013. This page generated on 2013-01-17T21:44:13Z



Roma: generating validators for the TEI

These pages will help you design your own TEI validator, as a DTD, RELAXNG or W3C Schema.

Create a new or upload existing customization

- ☐ Build schema (Create a new customisation by adding elements and modules to the smallest recommended schema)
- ☐ Reduce schema (Create a new customization by removing elements and modules from the largest possible schema)
- ☒ Create customization from template
- ☐ Open existing customization

Start

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Sanity check written by Ioan Bernevig. Please direct queries to the [TEI @ Oxford](#) project.

- Build up : ajouter ce qu'il vous faut à un schéma minimale
- Reduce : prenez tout ce qui existe et enlever ce qui ne vous interesse pas
- Utilisez un template ou un schema prédéfini
- Retravailler une personnalisation déjà faite

Contraintes schématron

- Une spécification d'élément peut aussi contenir un élément `<constraintSpec>`, rassemblant des règles exprimées en ISO schématron

```
<elementSpec xmlns:s="http://purl.oclc.org/dsdl/schématron"
ident="div" module="teisttructure" mode="change">
  <constraintSpec ident="cartoon" scheme="isoschématron">
    <constraint> <s:assert test="@type='cartoon' and ../tei:graphic">
      un cartoon doit inclure une graphie </s:assert> </constraint>
    </constraintSpec> </elementSpec>
```

Attention :

- le traitement de ces règles nécessite une étape supplémentaire dans la validation des documents
- ces règles ne sont pas forcément traitables par toute langue de schéma

Etre conforme à la TEI veut dire quoi?

- être honnet : Les éléments XML qui se déclarent comme appartenant au namespace TEI doivent respecter les définitions TEI de ces elements
- être explicite : Pour valider un document TEI, un ODD est fortement conseillé, parce que cela mettra en évidence toutes les modifications effectuées
- Tout document valide TEI est valide par rapport au schéma TEI-ALL
- Un document TEI peut contenir des elements de plusieurs namespaces, mais les elements dans le TEI namespace doivent respecter la semantique definie pour ces elements par la TEI

L'objet de ces règles est de faciliter le "blind interchange" des documents.

Recherche et analyse des documents TEI XML

- Au delà de l'édition ou de l'affichage des textes...
- Systèmes de indexation et de recherche: BD XML
- Systèmes d'intégration et d'accès "non-linéaire" aux ressources XML

XML (et donc TEI) comme composant du Web des Données

Les technologies incontournables

- XHTML : standard XML pour les pages Web
- CSS : standard pour les règles de mise en forme des pages Web
- JavaScript et son sous-ensemble AJAX : langage de script pour animer des pages Web
- PHP : langage de script pour création dynamique des pages web
- bases de données XML (par ex. eXist, BaseX)
- XQuery : langage de requête pour documents XML
- moteurs de recherche (par ex. Lucene)
- serveurs HTTP comme Apache
- etc.

Où sont les outils TEI-XML?

- La TEI elle-même ne vous fournit ni usine à gaz, ni boîte d'outils..
- Dès le début, elle essaie de se distinguer nettement de la production des outils, pour mieux garantir son indépendance
- Les **Guidelines** sont conçus comme expression concrète d'un modèle abstraite des objets -- pour la plupart textuels -- qui sont d'intérêt scientifique dans les SHS...
- Mais c'est réservé aux communautés d'utilisateurs de se décider comment traiter ces objets, et donc de construire les outils pour en profiter.

OK, mais quand même...

Les TEI Guidelines s'expriment en XML, comme toute application TEI

- en principe, alors, n'importe quel logiciel conforme au norme XML nous suffirait
- mais il faut le construire!

La TEI a construit, pour ses propres besoins, un ensemble de feuilles de style XSLT, qui permet:

- la génération et la documentation des schémas TEI personnalisés ("ODD")
- la visualisation des documents TEI, à travers les formats actuellement les plus répandus (XHTML, PDF, Word, Open Office, ePub...)
- une architecture générique pour la transformation entre documents TEI et d'autres formats (Word, OO, Docbook etc.)

Les Stylesheets de Sebastian Rahtz

(Il faut trouver un nom plus sexy...)

Disponible sous licence ouverte LGPL, ils se trouvent partout:

- emballés avec oXygen, par défaut
- téléchargeables en form de "package" pour Debian/Ubuntu, sur sourceforge
- à l'interieur du web service EGE (on va voir cela)
- soujacent plusieurs applications indépendentes de la TEI

Pourquoi s'en servir

- ils proposent des solutions aux difficultés que vous n'auriez pas (peut être) encore envisagés
- mieux vaut participer à un projet open source que de réinventer (plusieurs) roues
- ils sont maintenus en tandem avec les Guidelines
- ils sont modulaires, extensibles, bien "architecturés"

Pourquoi ne pas s'en servir

Ils ont des objectifs génériques, mais avec des limites!

- Ils ne traitent pas tout élément TEI!
- Ils ne traitent pas tous les valeurs possibles des attributs TEI (surtout ceux qui ne sont pas spécifiés par la TEI)

Pour en profiter, il faut:

- savoir se débrouiller en XSLT (un peu)
- ajouter une couche de personnalisation propre à son propre projet (probablement).

OxGarage

Un outil web qui permet une gamme de transformations à double sens

- OpenOffice XML (ODT) vers TEI (et l'inverse)
- Word 2007 XML (DOCX) vers TEI (et l'inverse)
- WordPerfect vers TEI (et l'inverse)
- Docbook vers TEI
- TEI vers ePub
- TEI P4 vers TEI P5
- ...

Disponible à <http://www.tei-c.org/ege-webclient/>

Open TEI

- La TEI n'est pas un projet de recherche mais une infrastructure de recherche
- Ses produits sont distribués sous licence GPL
- Sa gestion et son évolution se font en publique, sur Sourceforge cf <http://tei.sf.net> et ses composants sont disponibles sous Debian Linux (etc.)
- La TEI est une initiative communautaire, gérée par et pour les membres de cette communauté.
- Inscription à la TEI reste et restera facultative – l'essentiel, c'est la participation.
- This means you!