

Il y a une vie après MARC, 19 novembre 2007 : journée d'étude en hommage à Pierre-Yves Duchemin

TEI : Text Encoding Initiative

BENOIT, Jean-Luc

BENOIT, Jean-Luc. TEI : Text Encoding Initiative. In *Il y a une vie après MARC : journée d'étude organisée en hommage à Pierre-Yves Duchemin, l'enssib à Villeurbanne, 19 novembre 2007* [en ligne]. Format PDF.

Disponible sur : <<http://www.enssib.fr/bibliotheque-numerique/notice-1335>>



Ce document est diffusé sous licence « **Creative Commons by-nc-nd** ». Cette licence signifie que le document est mis à disposition selon le contrat **Paternité-Pas d'Utilisation Commerciale-Pas de Modification**, disponible en ligne à l'adresse <http://creativecommons.org/licenses/by-nc-nd/2.0/fr/> Il est ainsi possible de reproduire, distribuer et communiquer cette création au public, à condition de le faire à titre gratuit, mais ni de le proposer à titre onéreux ni le modifier sans le consentement explicite de l'auteur.

L'ensemble des documents mis en ligne par l'enssib sont accessibles à partir du site : <http://www.enssib.fr/bibliotheque-numerique/>



ANALYSE ET TRAITEMENT
INFORMATIQUE
DE LA LANGUE FRANÇAISE

Unité mixte de recherche 7118



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Nancy-Université



< Text Encoding Initiative >

Il y a une vie après MARK, Hommage à Pierre-Yves Duchemin

Lundi 19 novembre , ENSSIB.

J-luc Benoit

benoit@atilf.fr

<http://www.atilf.fr>

Le projet initié par PYD donne ses fruits

TEI Internationalization proposal

- Traduire les chapitres des Guidelines.
- Traduire la définition des éléments et leur description.
- Traduire le nom des éléments (<addrLine> → <ligneAdresse>).
- Introduire des exemples en français



- **Commission : CN 357 Modélisation, production et accès aux documents : AFNOR/CG 46/CN 357.**
 - Groupe d'experts : GE 8 Encodage des textes.



Le projet initié par PYD donne ses fruits

```
<elementSpec xmlns="http://www.tei-c.org/ns/1.0" module="core" xml:id="L" usage="mwa" ident="l">
  <equiv/>
  <gloss>verse line</gloss>
  <gloss version="2007-06-12" xml:lang="fr">ligne de vers</gloss>
  <gloss version="2007-05-04" xml:lang="es">verso</gloss>
  <gloss version="2007-05-02" xml:lang="zh-tw">詩行</gloss>
  <gloss version="2007-01-21" xml:lang="it">verso</gloss>
  <gloss version="2006-10-28" xml:lang="ja">詩の1行</gloss>
  <desc>contains a single, possibly incomplete, line of verse.</desc>
  <desc version="2007-06-12" xml:lang="fr">contient une seule ligne de vers, éventuellement
  incomplète</desc>
  <desc version="2007-05-04" xml:lang="es">contiene un único verso, posiblemente incompleto.</desc>
  <desc version="2007-05-02" xml:lang="zh-tw">包含詩文的一行，也許是不完整的詩行。</desc>
  <desc version="2007-01-21" xml:lang="it">contiene un singolo verso, anche incompleto.</desc>
  <desc version="2006-10-28" xml:lang="ja">完全な1行ではない場合もある。</desc>
  <classes>
```

Le projet initié par PYD donne ses fruits

VI GEORGES ET JEANNE

Moi qu'un petit enfant rend tout à fait stupide,
J'en ai deux; George et Jeanne; et je prends l'un pour guide
Et l'autre pour lumière, et j'accours à leur voix,
Vu que George a deux ans et que Jeanne a dix mois.

Le projet initié par PYD donne ses fruits

(division du texte)

```
<div type="poésie">
  <head>VI </head>
  <head> GEORGES ET JEANNE </head>
  <lg >
    <l>Moi qu'un petit enfant rend tout à fait stupide, </l>
    <l>J'en ai deux; George et Jeanne; et je prends l'un pour guide </l>
    <l>Et l'autre pour lumière, et j'accours à leur voix, </l>
    <l>Vu que George a deux ans et que Jeanne a dix mois. </l>
  </lg >
</div>
```

(ligne de vers)

- met (metrical structure, conventional) Contient un encodage spécifique utilisateur pour la structure métrique conventionnelle de l'élément.
- n
- org
- part
- real
- rend
- rendition

Plan



< Text Encoding Initiative >

- **Histoire.**
- **Définition.**
- **Usage.**
- **Délivrables.**
- **Description.**
- **Modus operandi.**
- **Héritage.**
- **Outils.**
- **C'est parti !**



Il était une fois,

il y a longtemps

dans une galaxie lointaine...

une histoire

- Il était une fois à Poukeespie...



1987: Vassar College Conference

une histoire

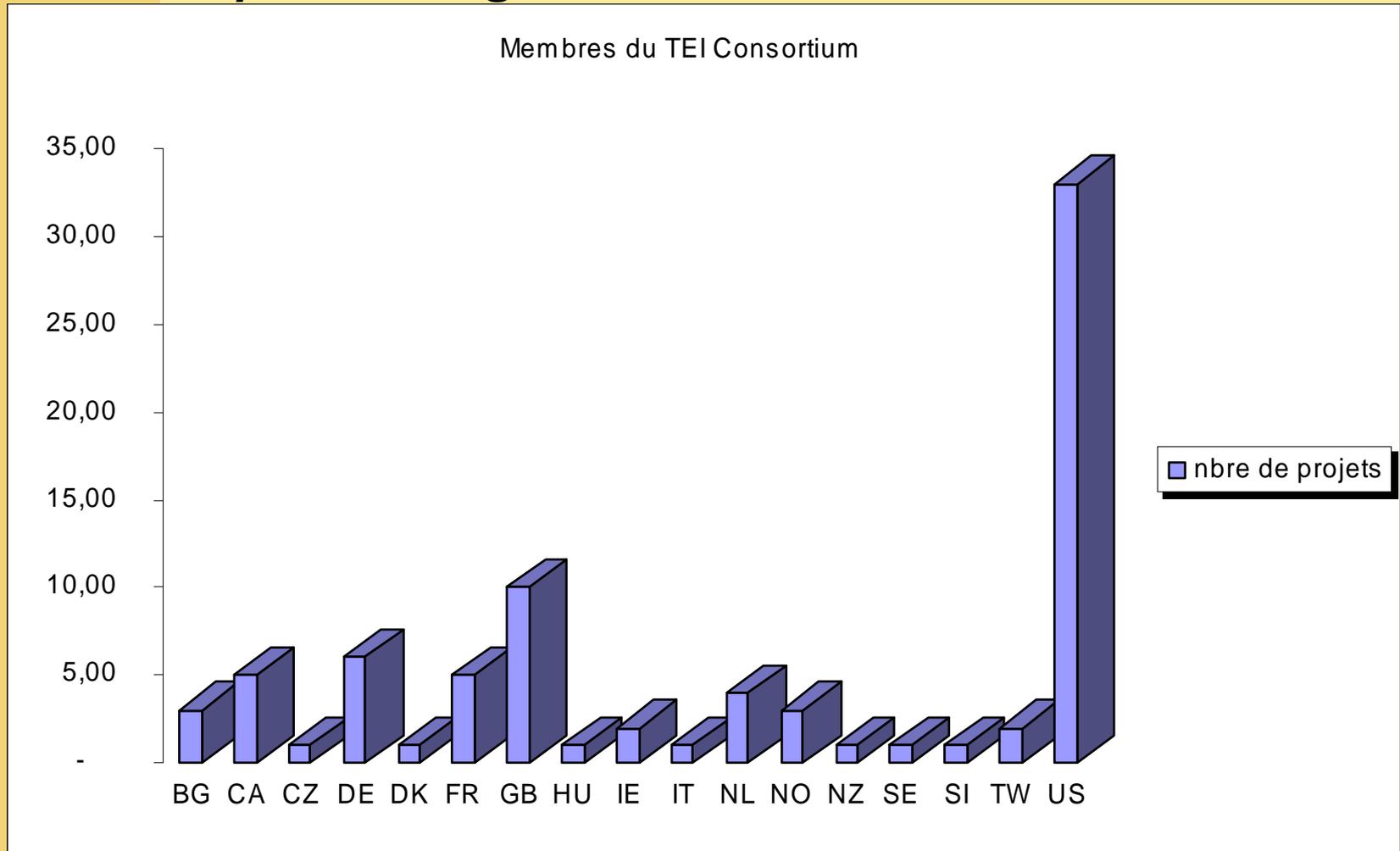


- *Novembre 1987, naissance à Poughkeepsie (New York).*
- *Janvier 1999, création du TEI Consortium.*
- *Aujourd'hui : une centaine de projets, quatre-vingt membres.*



Une histoire

Aujourd'hui : une centaine de projets, quatre-vingt membres.





Une histoire

- **communauté internationale de recherche.**
 - surtout dans sciences humaines, et en linguistique.
- **effort international.**
 - financement des États-Unis, de l'Union Européenne et du Canada).
- **parrainage par trois associations érudites.**
 - ACH (Association for Computers and the Humanities).
 - ALLC (l'Association for Computational Linguistics).
 - ACL (Association for Literary and Linguistic Computing).



TEI P5 Version 1.0

Enfin !

A Major New Release
of the TEI Guidelines

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>



< Text Encoding Initiative >



TEI P5 Version 1.0

- **Changements.**
 - Description des manuscrits, le multimédia et la visualisation graphique.
 - Amélioration dans la représentation des.
 - Abréviations.
 - Corrections.
 - Révision complète des pointeurs et des liens.
 - Traitement automatique des dates conforme aux normes ISO et W3C.
 - Accès à des versions traduites des Guidelines.

TEI P5 Version 1.0

P5 Fascicule

[Skip links](#)

[TEI P5 Home](#)

Text Encoding Initiative Consortium

TEI P5 Home

5 En-tête TEI

[1.1 Organisation de l'En-tête TEI](#)

[1.2 Description bibliographique du fichier](#)

[1.3 Description de l'encodage](#)

[1.4 Description du profil](#)

[1.5 Description des révisions](#)

[1.6 En-têtes minimaux et recommandés](#)

[1.7 Note pour les catalogueurs en bibliothèque](#)

[5 En-tête TEI](#)

[2. Class catalogue](#)

[3. Pattern catalogue](#)

[4. Element catalogue](#)

5 En-tête TEI

Ce chapitre traite des problèmes de description d'une oeuvre encodée afin que le texte lui-même, sa source, son balisage et ses révisions soient correctement documentés. Cette documentation est également nécessaire pour les chercheurs qui travaillent sur les textes, pour les catalogueurs dans les bibliothèques et les archives et pour les logiciels qui traitent ces textes. Ces descriptions et déclarations fournissent ensemble un substitut électronique à la page de titre d'une oeuvre imprimée. Elles constituent un équivalent au contenu des guides d'encodage ou des manuels d'introduction qui accompagnent généralement les jeux de données électroniques.

Tout texte conforme à la TEI doit comporter en préambule ce jeu de description, balisé conformément aux principes directeurs figurant dans ce chapitre. Cet ensemble de données est appelé En-tête TEI *header*, balisé `<teiHeader>` et comporte quatre parties principales :

- une description bibliographique du fichier *file description*, balisée `<fileDesc>`, qui contient une description bibliographique complète du fichier informatique lui-même, dont un utilisateur du texte pourra extraire une citation bibliographique, et qu'un bibliothécaire ou un archiviste pourra utiliser dans un catalogue. Le terme « fichier informatique » *computer file* doit être compris comme faisant référence à l'intégralité du document décrit dans l'en-tête, même s'il est conservé dans plusieurs fichiers distincts. La description du fichier contient également des informations sur la ou les source(s) à l'origine du document électronique. Les éléments TEI utilisés pour encoder la description de fichier sont décrits à la section 5.2 Description bibliographique du fichier [1.2 Description bibliographique du fichier](#).
- une description de l'encodage *encoding description*, balisée `<encodingDesc>`, qui décrit les relations existant entre le texte électronique et la ou les source(s) dont il dérive. Elle permet une description détaillée des éléments permettant de savoir si (et comment) le texte a été normalisé pendant sa transcription, comment l'encodeur a résolu les ambiguïtés de la source, quels niveaux d'encodage ou d'analyse ont été appliqués, etc. Les éléments TEI utilisés pour encoder la description de l'encodage sont décrits à la section 5.3 Description de l'encodage [1.3 Description de l'encodage](#).
- un profil de texte *text profile*, balisé `<profileDesc>`, qui contient des informations contextuelles et de classification sur le texte, telles que son sujet, le contexte dans lequel il a été produit, les individus mentionnés dans le texte ou ayant participé à sa création, etc. Cet élément est d'un intérêt particulier dans le cas de textes composites fortement structurés, tels que des corpus ou des collections linguistiques pour lesquels il est fortement souhaitable d'appliquer un vocabulaire descriptif contrôlé ou d'extraire du corps du texte des termes relatifs au type de texte ou à son origine. Le profil de texte peut toutefois être utile dans tous les cas de traitement automatique de texte. Les éléments TEI utilisés pour encoder le profil de texte sont décrits à la section 5.4 Description du profil [1.4 Description du profil](#).
- un historique des révisions *revision history*, balisé `<revisionDesc>`, qui permet à l'encodeur de fournir un historique des modifications effectuées au cours de l'élaboration du texte électronique. L'historique des révisions est important pour le suivi des versions *version control* et pour répondre aux questions relatives à l'évolution d'un fichier. Les éléments TEI



TEI P5 Version 1.0

Repose sur le langage de Schéma XML

- **Effectue des validations plus rigoureuses.**
- **Utilise des éléments empruntés à d'autres vocabulaires reposant sur XML comme MathML ou Docbook dans un document TEI.**
- **Implémente des documents TEI dans d'autres documents XML comme METS et MODS.**
- **Crée, gère, documente plus facilement des personnalisations générées par les utilisateurs.**



TEI P5 Version 1.0

Les caractéristiques qui ont fait de la TEI le standard d'encodage préféré des 20 dernières années est conservé.

- **L'accès libre et gratuit sur le Web.**
- **Une architecture modulaire qui peut être facilement être paramétrable.**
- **Un nombre très importants d'éléments capables de nuancer l'encodage de manière très fine et qui est capable de s'adapter à la presque totalité des textes.**

TEI : définition.



- *Text Encoding Initiative.*
- *Encoding and Interchange.*
- *Groupe d'initiative pour un encodage normalisé.*



{ Digression : encodage ? 1 }

Un format d'échange s'exprime dans un encodage.

Pourquoi encoder ? Que veut-on veut encoder ?

- **L'encodage exprime un aspect quelconque d'un document.**
- **tout encodage est donc une interprétation.**
- **aucun langage d'encodage ne peut donc prétendre être ni complet ni exhaustif.**

{ Digression : encodage ? ² }

Un langage d'encodage sert à...

- **Spécifie les caractéristiques d'un texte.**
- **Rend explicite la/les structures aperçue/s dans un texte.**
- **Spécifie les méta-informations, renseignements contextuels etc.**



A quoi sert la TEI ?

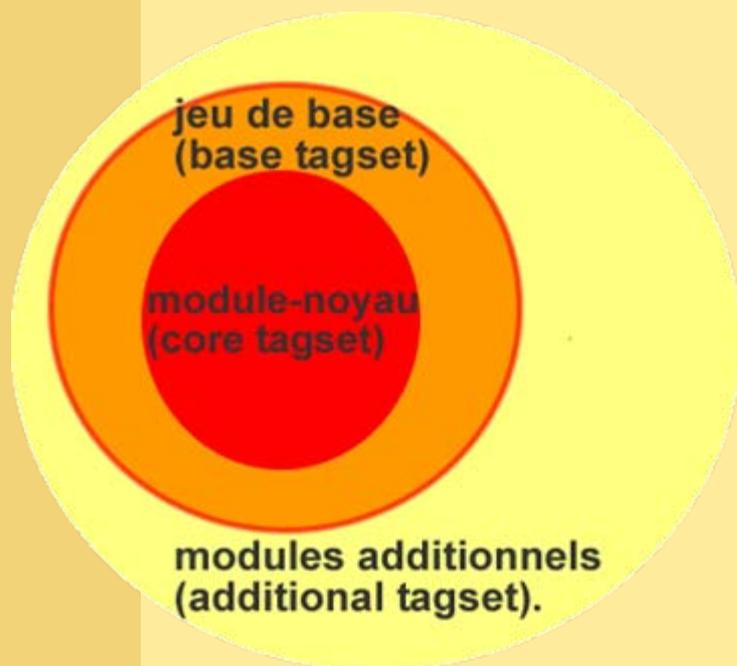
- ***Faciliter la création, l'échange, et l'intégration des données textuelles informatisées.***
 - toute sorte de texte.
 - toutes les langues.
 - toute provenance temporelle ou culturelle.
- ***La TEI doit servir également à aider...***
 - les *débutants*, cherchant des solutions bien connues et consensuelles.
 - les *experts*, cherchant à créer de nouvelles solutions.



Les “délivrables” de la TEI ?

- **Un ensemble cohérent de recommandations sur l’encodage, fondé sur la pratique actuelle.**
 - C’est cet ensemble que P-Y. D. avait entrepris de traduire.
- **un système extensible, modulaire, polymorphe.**
 - C’est sur cet ensemble qu’a porté notre traduction depuis l’été.
- **une documentation extensive.**
- **Manuels, documentations de référence (P3, P4, P5 et...).**

Description



Un système de balisage organisé selon trois types d'éléments.

- *le noyau.*
- *les balises et éléments propres à des disciplines.*
- *modules additionnels.*



Description

module-noyau

- Header(54 éléments).
 - ✓ ... biblFull, ... encodingDesc, extent, fileDesc, ...,profileDesc, projectDesc, ... tagUsage, tagsDecl, taxonomy, teiHeader, textClass, titleStmt, ...

- Core (78 éléments).
 - ✓ abbr, ..., author,...cit, corr, ... hi, imprint, index, item, l, label, lb, lg, list, listBibl, measure, measureGrp, meeting, mentioned, milestone, monogr, ...

- Textstructure (33 éléments).
 - ✓ TEI, ..., back, body, ..., closer, dateline, div, ... docAuthor, docDate, ..., front, ... text, titlePage, titlePart, trailer...

Le TEI Header

- **description bibliographique normalisée.**
 - du document électronique (titre, responsabilité , maison d'édition, source...).
 - de son encodage (éléments présents, codes internes...).
 - de sa classification (sujets, genres...).
 - de son état de revision.
- **facilite la découverte des ressources sur réseau et dans les bases de données.**

	Include	Exclude	Name	Description	Attributes
applInfo	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="applInfo"/>	records information about an application which has edited the TEI file.	Change attributes
application	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="application"/>	provides information about an application which has acted upon the document.	Change attributes
authority	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="authority"/>	fournit le nom d'une personne ou d'une institution responsable de la mise à disposition d'un texte électronique, autre qu'un éditeur ou un distributeur.	Change attributes
availability	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="availability"/>	fournit des informations sur la disponibilité du texte, par exemple toutes restrictions sur son usage ou sa diffusion, son statut de copyright, etc.	Change attributes
biblFull	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="biblFull"/>	contient une citation bibliographique entièrement structurée, dans laquelle tous les composants de la description du fichier TEI sont présents.	Change attributes
cRefPattern	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="cRefPattern"/>	précise une expression et un modèle de remplacement pour convertir une référence canonique en une URI	Change attributes
catDesc	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="catDesc"/>	décrit une certaine catégorie à l'intérieur d'une taxinomie ou d'une typologie de texte, sous forme d'une brève description en prose suivie ou en termes de paramètres contextuels utilisés dans l'élément textDesc(Description du texte)	Change attributes
catRef	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="catRef"/>	spécifie une ou plusieurs catégories définies dans une taxinomie ou une typologie de textes.	Change attributes
category	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="category"/>	contient une catégorie descriptive individuelle, éventuellement intégrée dans une catégorie de niveau supérieur, à l'intérieur d'une taxinomie définie par l'utilisateur.	Change attributes
change	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="change"/>	résume une modification ou une correction apportée à une version particulière d'un texte électronique partagé entre plusieurs chercheurs.	Change attributes
classCode	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="classCode"/>	contient l'indice attribué à ce texte en référence à un système standard de classification.	Change attributes
classDecl	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="classDecl"/>	contient une ou plusieurs taxinomies définissant tout code classificateur utilisé ailleurs dans le texte.	Change attributes
correction	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="correction"/>	établit comment et dans quelles circonstances des corrections ont été apportées au texte.	Change attributes
creation	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="creation"/>	contient des informations concernant la création d'un texte.	Change attributes
distributor	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="distributor"/>	fournit le nom d'une personne ou d'une institution responsable de la diffusion d'un texte	Change attributes
edition	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="edition"/>	Décrit les particularités de l'édition d'un texte.	Change attributes
editionStmnt	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="editionStmnt"/>	regroupe les informations relatives à l'édition d'un texte.	Change attributes
editorialDecl	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="editorialDecl"/>	fournit des détails sur les pratiques et principes éditoriaux appliqués lors de l'encodage d'un texte.	Change attributes
encodingDesc	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="encodingDesc"/>	documente la relation entre un texte électronique et la ou les sources dont il dérive.	Change attributes

Description

```
<sourceDesc>
  <biblStruct>
    <monogr>
      <author>M. Marivaux</author>
      <title>Théâtre complet</title>
      <editor>Édition établie par <name>Frédéric Deloffre</name> et
        <name>Françoise Rubellin</name>
      </editor>
      <imprint>
        <publisher>Garnier</publisher>
        <pubPlace>Paris</pubPlace>
        <date>2000</date>
      </imprint>
    </monogr>
    <idno type="ISBN">2253132535</idno>
    <idno type="FRANTEXT">S175</idno>
  </biblStruct>
</sourceDesc>
```

TEI Header (extrait)

Description

Core TEI

Description
module-noyau
- Header (54 éléments). ✓ ... biblFull, ... encodingDesc, extent, fileDesc, ... profileDesc, projectDesc, ... tagUsage, tagsDecl, taxonomy, teiHeader, textClass, titleStmt, ...
- Core (78 éléments). ✓ abbr, ..., author, ... cit, corr, ... hi, imprint, index, item, l, label, lb, lg, list, listBibl, measure, measureGrp, meeting, mentioned, milestone, monogr, ...
- Textstructure (33 éléments). ✓ TEI, ..., back, body, ..., closer, dateline, div, ... docAuthor, docDate, ..., front, ... text, titlePage, titlePart, trailer...

- ensemble des éléments et des attributs requis pour tous les genres de documents.
- appartenant au module noyau mais n'étant pas un élément structurant.
 - éléments descriptifs (noms, dates, mises en relief...)
 - modifications éditoriales
 - liens, références...

List of elements in module: core

	Include	Exclude	Name	Description	Attributes
abbr	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="abbr"/>	? contient une abréviation quelconque	Change attributes
add	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="add"/>	? contient des lettres, des mots ou des phrases insérés dans le text par un auteur, un copiste, un annotateur ou correcteur	Change attributes
addrLine	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="addrLine"/>	? contient une ligne d'une adresse postale	Change attributes
address	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="address"/>	? contient une adresse, postale ou autre, par exemple l'adresse d'un éditeur, d'une organisation ou d'un individu	Change attributes
analytic	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="analytic"/>	? contient des éléments de description bibliographique d'une ressource (par exemple un poème ou un article de revue) publié à l'intérieur d'une monographie ou d'une ressource continue et non pas publié de façon indépendante	Change attributes
author	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="author"/>	? dans une référence bibliographique contient le nom de la (ou des) personne(s) physique(s) ou collectivité(s), auteur(s) d'une oeuvre ; la première mention de responsabilité pour tout object bibliographique	Change attributes
bibl	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="bibl"/>	? contient une citation bibliographique faiblement structurée dans laquelle les sous-composants peuvent ou non être explicitement balisés	Change attributes
biblScope	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="biblScope"/>	? définit l'étendue d'une référence bibliographique, par exemple sous forme d'une liste de numéros de pages, ou le nom d'une subdivision d'une oeuvre plus large	Change attributes
biblStruct	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="biblStruct"/>	? contient une citation bibliographique dans laquelle seuls des sous-éléments bibliographiques apparaissent et ce, dans un ordre déterminé	Change attributes
binaryObject	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="binaryObject"/>	? fournit des données binaires encodées qui représentent une image ou un autre objet insérés dans le texte ou un autre objet	Change attributes
cb	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="cb"/>	? marque la limite entre une colonne d'un texte et la colonne suivante dans un système de référence normalisé	Change attributes
choice	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="choice"/>	? regroupe un nombre de balisages alternatifs pour le même point d'un texte	Change attributes
cit	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="cit"/>	? citation provenant d'un autre document avec la référence bibliographique de sa source	Change attributes
corr	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="corr"/>	? contient la forme correcte d'un passage apparemment erroné dans l'exemplaire du texte	Change attributes

Description

```
<cit>
<quote>
<hi rend="I"> On ne sait plus qui sont ces passants de l'automne</hi>
Mais sais-tu qui tu es ? Mais sais-tu d'où tu viens ?</hi>
Es-tu vraiment toi-même ? Quelqu'un qui serait toi ?</hi>
</b>
</b>
<hi rend="I"> Es-tu le souvenir d'un autre que tu fus ?</hi>
Es-tu sûr d'être là ? Et si tu n'es pas là</hi>
qui est cet étranger qui parle avec ta voix ?</hi>
</quote>
<bibl>Claude Roy</bibl>
</cit>
```

```
<author>Bernard-Marie Koltès</author>
<title>Quai ouest</title>
<date>avril 2001</date>
<publisher>Les Editions de Minuit</publisher>
<pubPlace>Paris</pubPlace>
```

Core TEI

Description

Description

module-noyau

- Header(54 éléments).
 - ✓ ... biblFull, ... encodingDesc, extent, fileDesc, ... profileDesc, projectDesc, ... tagUsage, tagsDecl, taxonomy, teiHeader, textClass, titleStmt, ...
- Core (78 éléments).
 - ✓ abbr, ..., author, ... cit, corr, ... hi, imprint, index, item, l, label, lb, lg, list, listBibl, measure, measureGrp, meeting, mentioned, milestone, monogr, ...
- Textstructure (33 éléments).
 - ✓ TEL, ..., back, body, ..., closer, dateline, div, ... docAuthor, docDate, ..., front, ... text, titlePage, titlePart, trailer...

- **Textstructure**

- éléments structuraux (divisions, paragraphes, vers, discours dramatiques, titres, notes, listes...)

	Include	Exclude	Name	Description	Attributes
TEI	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="TEI"/>	 contient un seul document, conforme à la TEI, qui comprend un en-tête TEI et un texte soit de façon isolée soit comme une partie d'un élément TEI corps	Change attributes
argument	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="argument"/>	 liste formalisée ou description en texte libre des sujets traités dans une subdivision d'un texte	Change attributes
back	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="back"/>	 contient tout ce qui suit le corps du texte : appendice, etc.	Change attributes
body	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="body"/>	 contient la totalité du corps d'un seul texte unitaire, à l'exclusion de toute partie liminaire	Change attributes
byline	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="byline"/>	 contient la mention de responsabilité principale pour une œuvre donnée sur sa page de titre ou bien au début ou à la fin de l'œuvre	Change attributes
closer	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="closer"/>	 regroupe formule de politesse, zone de date et autres expressions semblables figurant comme groupe à la fin d'une division, en particulier d'une lettre	Change attributes
dateline	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="dateline"/>	 contient une brève description des lieux, date, heure, etc. de production d'une lettre, d'un article de journal ou d'un autre texte qui lui sont rattachés au début ou à la fin comme en-tête ou annonce de fin	Change attributes
div	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div"/>	 contient une subdivision d'une partie liminaire ou du corps d'un texte	Change attributes
div1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div1"/>	 contient une subdivision au premier niveau d'une partie liminaire ou du corps d'un texte	Change attributes
div2	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div2"/>	 contient une subdivision au deuxième niveau des parties liminaires ou du corps d'un texte	Change attributes
div3	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div3"/>	 contient une subdivision au troisième niveau d'une partie liminaire ou du corps d'un texte	Change attributes
div4	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div4"/>	 contient une subdivision au quatrième niveau d'une partie liminaire ou du corps d'un texte	Change attributes
div5	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div5"/>	 contient une subdivision au cinquième niveau d'une partie liminaire ou du corps d'un texte	Change attributes
div6	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div6"/>	 contient une subdivision au sixième niveau d'une partie liminaire ou du corps d'un texte	Change attributes
div7	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="div7"/>	 contient la plus petite subdivision possible d'une partie liminaire ou du corps d'un texte mais plus grande qu'un paragraphe	Change attributes
docAuthor	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="docAuthor"/>	 contient le nom de l'auteur du document tel qu'il est donné sur la page de titre (ce nom est le plus souvent contenu dans une zone de signature)	Change attributes
docDate	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="docDate"/>	 contient la date d'un document telle qu'elle est généralement donnée sur une page de titre	Change attributes
docEdition	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="docEdition"/>	 contient une mention d'édition telle qu'elle figure sur la page de titre d'un document	Change attributes
docImprint	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="docImprint"/>	 contient la mention de dénomination commerciale d'éditeur (lieu et date de publication, nom de l'éditeur), telle qu'elle est	Change attributes

Textstructure

```
<text>
<front>
  <pb n="2"/>
  <div type="preface">
    <head>PR&#201;FACE</head>
    <p>Tant qu'il existera, par le fait des lois et des moeurs, une damnation sociale
      créant artificiellement, en pleine civilisation, des enfers, et compliquant
      d'une fatalité humaine la destinée qui est divine ; tant que les trois problèmes
      du siècle, la dégradation de l'homme par le prolétariat, la déchéance de la
      femme par la faim, l'atrophie de l'enfant par la nuit, ne seront pas résolus ;
      tant que, dans de certaines régions, l'asphyxie sociale sera possible ; en
      d'autres termes, et à un point de vue plus étendu encore, tant qu'il y aura sur
      la terre ignorance et misère, des livres de la nature de celui-ci pourront ne
      pas être inutiles.</p>
    <closer>
      <dateline>
        <name>Hauteville-House,</name>
        <date>1862</date>.</dateline>
      </closer>
    <pb n="3"/>
  </div>
</front>
<!-- fin de front -->
<body>
  <div type="partie">
    <head>première partie</head>
    <head>Fantine</head>
    <div type="livre premier">
      <head>Livre premier</head>
      <head>Un juste</head>
      <div type="sous-partie">
        <head>I</head>
        <head>
          <abbr>M.</abbr>MYRIEL</head>
        <p>EN 1815, <abbr>M.</abbr>Charles-François-Bienvenu Myriel était évêque de Digne
          <!-- [...] --></p>
      </div>
    </div>
  </div>
</body>
</text>
```



Description

module-base

- **définit les composants fondamentaux des documents.**
- **bases définies :**
 - prose, poésie, théâtre.
 - transcription du discours oral.
 - dictionnaires, terminologies.
- **on peut aussi mélanger les bases.**
 - (avec précaution).

module: dictionaries

	Include	Exclude	Name	Description	Attributes
case	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="case"/>	contient des informations sur le cas grammatical présenté par le dictionnaire pour une forme donnée.	Change attributes
colloc	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="colloc"/>	contient une collocation de l'entrée	Change attributes
def	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="def"/>	contient le texte de la définition contenue dans l'entrée	Change attributes
dictScrap	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="dictScrap"/>	contient la partie d'une entrée de dictionnaire dans laquelle d'autres éléments de niveau "expression" sont librement associés.	Change attributes
entry	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="entry"/>	contient une entrée structurée de dictionnaire.	Change attributes
entryFree	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="entryFree"/>	contient une entrée de dictionnaire qui ne se conforme pas nécessairement aux contraintes imposées par l'élément entry.	Change attributes
etym	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="etym"/>	dans une entrée, contient les informations étymologique	Change attributes
form	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="form"/>	regroupe toutes les informations relatives à la morphologie et à la prononciation d'une entrée	Change attributes
gen	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="gen"/>	identifie le genre morphologique d'un élément lexical, tel qu'il est donné par le dictionnaire.	Change attributes
gram	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="gram"/>	contient de l'information grammaticale relative à un terme, un mot, ou une forme dans une entrée de dictionnaire ou dans un fichier de données terminologiques.	Change attributes
gramGrp	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="gramGrp"/>	regroupe des informations morphosyntaxiques sur un item lexical, par exemple Partie du discours pos, Genre gen, Nombre number, Cas case, ou Classe flexionnelle iType.	Change attributes
hom	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="hom"/>	regroupe les informations relatives à un homographe dans une entrée.	Change attributes
hyph	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="hyph"/>	contient une entrée avec un trait d'union ou des informations relatives à l'usage du trait d'union dans une autre forme.	Change attributes
iType	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="iType"/>	indique la classe flexionnelle à laquelle appartient un item lexical.	Change attributes
lang	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="lang"/>	nom de la langue mentionnée dans une section relative à l'étymologie ou à un autre point linguistique	Change attributes
lbl	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="lbl"/>	étiquette pour la forme d'un mot, pour un exemple, pour une traduction, ou pour tout autre type d'information, par exemple "abréviation pour", "contraction de", "littéralement", "approximativement", "synonymes", etc.	Change attributes
mood	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="mood"/>	contient des informations sur le mode grammatical des verbes (par exemple l'indicatif, le subjonctif, l'impératif)	Change attributes
number	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="number"/>	indique le nombre grammatical associé à une forme, tel qu'elle est donnée par le dictionnaire.	Change attributes

Codage dictionnaire

<entry>

<form> NAïADE</form>

<gramGrp>, <pos>s. </pos><gen>f.</gen>

</gramGrp>

<sense>Fausse Divinité, que les Païens croyoient présider aux fontaines, & aux rivières. <trans xml:lang="la">

<hi rend="italique">Naias</hi>, <hi rend="italique">nager.</hi>

</trans>

<eg> Strabon dit, L. X. que les Naïades étoient des Prêtresses de Bacchus. Voyez Natal. Comes, L. V. c. 13.

Naïades, poursuit-il, Déesses des Fontaines,

Qui pour la ville avez quitté les plaines,

Venez au secours de Santeuil,

Il a chanté pour vous ; de son ame affligée,

N'adoucirez-vous point le deuil ?

Trad. du P. Commire.</eg>

<etym>Ce mot vient du Grec<lang>

<hi rend="italique">

>ὐἀἐἰὐ</hi>

<!--naein-->

</lang> , couler, ou

<hi>ὐἀἰἐἰὐ</hi>

<!--naiein-->, qui signifie <hi rend="italique">demeurer</hi>. Les

Poètes font souvent mention des <hi rend="italique">

>Naïades</hi>. Voyez Vossius, De Idolol. L. V. c.

80.</etym></sense>

</entry>

NAIADES, s. f. Fausse Divinité que les Païens croyoient présider aux fontaines & aux rivières. *Naias, nager*. Strabon dit L. X. que les *Naiades* étoient des Prêtresses de Bacchus. Voyez *Natal. Comes* L. V. C. 13.

*Naiades, poursuit-il, Déesses des Fontaines,
Qui pour la ville avez quitté les plaines,
Venez au secours de Santeuil,
Il a chanté pour vous ; de son ame affligée,
N'adoucirez-vous point le deuil ?*

TRAD. DU P. COMMIRE.

Ce mot vient du Grec *ναειν*, couler, *ναειν*, qui signifie *demeurer*. Les Poètes font souvent mention des *Naiades*.

NAÏADE, s. f. Fausse Divinité, que les Païens croyoient présider aux fontaines, & aux

rivières. *Naias, nager*. Strabon dit, L. X. que les *Naiades* étoient des Prêtresses de Bacchus. Voyez *Natal. Comes*, L. V. c. 13.

Naiades, poursuit-il, Déesses des Fontaines,

Qui pour la ville avez quitté les plaines,

Venez au secours de Santeuil,

Il a chanté pour vous ; de son ame affligée,

N'adoucirez-vous point le deuil ?

Trad. du P. Commire.

Ce mot vient du Grec (GREC), couler, ou (GREC), qui signifie *demeurer*. Les Poètes font

souvent mention des *Naiades*. Voyez *Vossius*, *De Idolol.* L. V. c. 80.



Description

Module additionnel

- **ajoute des éléments ou attributs spécialisés.**
 - linking (liens, jalons, segmentation.).
 - analyse et interprétation,
 - structures de traits (feature structures).
 - manuscrits et sources primaires.
 - appareil critique.
 - études onomastiques.
 - formules, tables, graphiques.
 - méta-informations pertinentes aux corpus.



Description

Attributs globaux

- **tout élément est membre de la classe «globale», et donc porte les attributs suivants:**
 - **id** (identification unique).
 - **n** (identification locale).
 - **lang** (langage, identification du WSD rélevant).
 - **rend** (rendition).
- **on peut étendre la classe globale en choisissant des jeux additionnels.**

Modus operandi

- identifier les *traits textuels significatifs* indépendamment de leur notation ou réalisation.
- se méfier des controverses, et des raffinements excessifs et des simplifications inutiles.
- chercher des solutions génériques.
- appui sur le contenu, plutôt que sur la présentation.
- un schéma libéral, peu normatif.
- le rasoir de Occam.
- une structuration fonctionnelle, avec plusieurs possibilités d'enrichissement.

Héritage de la TEI

- **Une façon de considérer ce que le texte est vraiment.**
- **Une codification des pratiques académiques courantes.**
- **Un ensemble d'a priori et de priorités partagées dans des perspectives numériques ... qui restent valables depuis plus de dix ans.**



Outils TEI

- **Consortium** <http://www.tei-c.org/index.xml>
- **Roma (successeur de Pizza-chef)** <http://tei.oucs.ox.ac.uk/Roma/>
- **Bibliothèque de feuilles de style.** <http://www.tei-c.org/Tools/Stylesheets/>
- **TEI wiki.** <http://www.tei-c.org.uk/wiki/index.php/Category:Tools>
- **Xaira.** <http://www.oucs.ox.ac.uk/rts/xaira/>.
- **Internationalisation de la TEI.** <http://www.tei-c.org/Tools/I18N/proposal.xml>

la traduction de la documentation de référence, celui des noms des éléments et des noms d'attribut ont été traduits dans cinq langues En français, en espagnol, en allemand, en chinois et en japonais.

Demain je me mets à la TEI !

- **Quels modules utiliserez-vous ?**
- **Comment allez-vous les personnaliser?**
- **Quelles contraintes supplémentaires sont nécessaires?**
- **Quels sont les logiciels que vous allez utiliser ?**