

TEI++: une formation approfondie pour les consortiums Cahier, Corpus Ecrits et IRCOM de la TGIR Corpus

Lou Burnard
avec la participation de : Alexandre Gefen, Lauranne Bertrand,
Alexei Lavrentiev, et Bertrand Gaiffe
Paris, novembre 2012

Introductions

- Formateurs et participants
- Modalités et organisation de la formation
- Un échantillon de projets TEI
- Principes basiques de l'encodage

Lundi 19 novembre

09h30 - 11h00 Introductions.

11h30 - 13h00 Session de rappels pratiques: éditer avec oXygen et notions de schéma.

14h00 - 15h30 Présentation du paysage TEI

16h00 - 17h30 Construction de schéma pour ses propres textes en se servant de Roma

Mardi 20 novembre

09h30 - 11h00 Présentation du langage ODD

11h30 - 13h00 Création d'un ODD à partir d'un corpus ;
modification d'un ODD

14h00 - 15h30 Présentation du TEI Header et considérations sur les
métadonnées

16h00 - 17h30 Création d'un header de manuscrit

Mercredi 21 novembre

09h30 - 11h00 Encodage des entités nommées avec analyses littéraires

11h30 - 13h00 Exercices "data mining" dans un corpus de documents historiques

14h00 - 15h30 Encodage TEI des corpus oraux avec analyses linguistiques

16h00 - 17h30 Exercice sur l'encodage d'un texte oral

Jeudi 22 novembre

- 09h30 - 11h00 Encodage TEI des sources manuscrites
- 11h30 - 13h00 Exercices sur l'encodage d'un manuscrit (a) transcription et (b) alignement texte/image
- 14h00 - 15h30 Stockage, affichage et analyse des corpus TEI ; discussion sur les outils
- 16h00 - 17h30 Experimentation avec quelques logiciels disponibles pour stockage, affichage et analyse des corpus TEI

Un florilège de projets TEI

- Un archive de textes littéraires
- Une édition numérique du correspondance d'un peintre célèbre
- Un corpus linguistique
- Une édition numérique de l'archive d'un scientifique du 17eme
- Un catalogue de manuscrits
- Un projet de recherche en édition génétique

- <http://wiki.tei-c.org/index.php/Samples>
- <http://www.tei-c.org/Activities/Projects/>

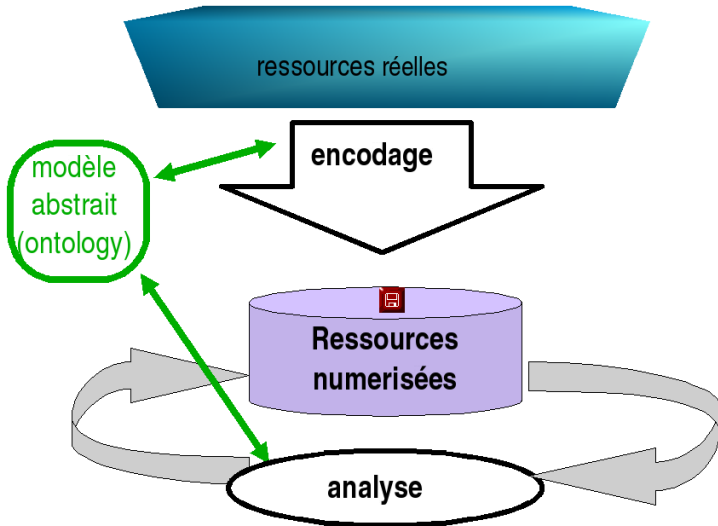
L'encodage : une technologie ancienne qui facilite et exprime nos lectures

- L'affichage des mots d'un texte n'est pas aleatoire !
- Pour la controller dans un texte numerique, le balisage est essentiel
- Mais il sert aussi a exprimer nos lectures prealables

L'encodage

- Un document numérique est plus qu'une séquence de caractères encodés!
- Un document est plus qu'une séquence de formes lexicales!
 - Il a une **structure** et une **signification**
 - Un document peut avoir plusieurs **lectures** variantes
 - La portée d'un document peut être **enrichie** par des annotations
- L'encodage explicite les lectures
- Sans explicitation, on ne peut rien traiter.

Qu'est-ce qu'on fait en numérisant un texte?



La TEI : un ensemble de définitions de balise

- le balisage XML n'a pas de sémantique
- La TEI lui apporte des règles d'usage et de signification
- Mais en même temps, la TEI doit permettre l'extension et la modification de ces règles...

La TEI s'adresse également aux débutants et aux experts