

# TEI à la carte (1)

Lou Burnard, MEET

2012-01



# Personnalisation des schémas TEI

- TEI au-dessous de la capote
- Construction d'un schéma TEI
- Personnalisation de la TEI
- Exploration des TEI Guidelines



## Reprise terminologique

- Le système TEI comprend plusieurs *modules*
- Chaque module comprend plusieurs *spécifications d'élément*
- Chaque spécification comprend:
  - un nom canonique (`<gi>`) en anglais pour l'élément et facultativement des noms équivalents en d'autre langues
  - une description canonique de sa fonction ou mode d'emploi (facultativement traduite en d'autre langues) fonction
  - une déclaration pour chacune des *classes* auxquelles il appartient
  - une définition pour chacun de ses *attributs*
  - une définition de son *modele de contenu*
  - des exemples d'usage, des notes, des liens
- une spécification de *schéma* TEI (`<schemaSpec>`) se fait par une sélection de modules ou d'éléments, avec (eventuellement) des modifications
- un document TEI qui contient une spécification de schema s'appelle (informellement!) un *ODD* (One Document Does it all)



## Qu'est-ce que c'est qu'un module?

- Une manière convenable de regrouper un ensemble de déclarations d'éléments
- D'habitude (mais pas forcément) associés avec champs de travail ou topos spécifique
- La plupart des chapitres de P5 fournissent la définition d'un seul module
- On crée un schéma TEI en sélectionnant des modules, et en supprimant ou en sélectionnant des éléments spécifiques de ces modules



# Liste des modules

## Nom des module

analysis  
certainty  
core  
corpus  
dictionaries  
drama  
figures  
gaiji  
header  
iso-fs  
linking  
msdescription  
namesdates  
nets  
spoken  
tagdocs  
tei  
textcrit  
textstructure  
transcr  
verse

## chapitre

Simple Analytic Mechanisms  
Certainty and Responsibility  
Elements Available in All TEI Documents  
Language Corpora  
Dictionaries  
Performance Texts  
Tables, Formulae, and Graphics  
Representation of Non-standard Characters and Glyphs  
The TEI Header  
Feature Structures  
Linking, Segmentation, and Alignment  
Manuscript Description  
Names, Dates, People, and Places  
Graphs, Networks, and Trees  
Transcriptions of Speech  
Documentation Éléments  
The TEI Infrastructure  
Critical Apparatus  
Default Text Structure  
Representation of Primary Sources  
Verse



## Comment choisir?

- On peut tout choisir! (pas vraiment une bonne idée)
- On peut partir d'une selection prédéfinie (TEI Lite, TEI Bare...)
- On peut faire artisanal -- selon les besoins spécifique de son projet

Dans ce dernier cas, il faut prendre conscience de toutes les possibilités disponibles...

**Roma** un logiciel en ligne qui simplifie cette procedure

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



## Un cas d'étude: la carte postale

Supposons qu'on a proposé un ANR pour la numérisation et transcription d'une collection gigantesque de cartes postales. Comment on va créer un schéma apte à gérer des centaines de choses comme celles-ci?

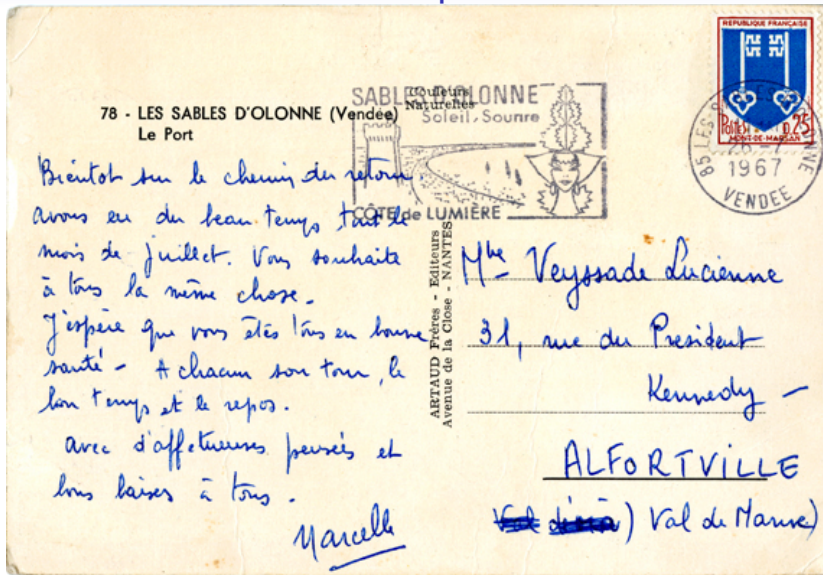


# Une carte postale

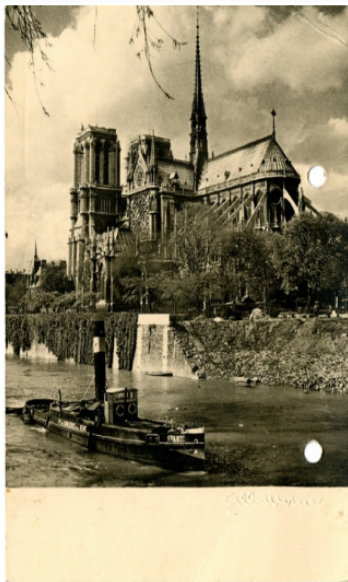




# Une carte postale



## Une autre carte postale



2/193 — PARIS — Notre-Dame. 26.7.53 — 1944

Mon amour

Je ne peux que penser à toi -  
 il m'est impossible de fixer  
 mon esprit, sur un autre  
 sujet (pratique) aussi important  
 soit-il - il n'y a pas plus  
 de Commerce - que de ministre -  
 d'avis, que de travail, et même  
 ma famille, père et mère, en  
 ont l'air surpris - Je rêve  
 littéralement - c'est fou. Je  
 t'en supplie laisse moi un peu  
 il faut penser à l'avenir, dont  
 tu parlais, heureux, riche, utile.  
 Je t'aime et... je ne pense qu'à toi (je)

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...



# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Quelles sont les éléments significatifs, qu'il faut absolument distinguer dans notre encodage?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- metadonnées à volonté
- ...

# Propositions

On va commencer en structurant chaque carte en divisions de type variés:

- **recto** : la surface qui porte l'image (normalement)
- **verso** : la surface qui porte (normalement) le message

Ensuite, on observe que on pourrait avoir indifféremment sur ces deux surfaces d'autres sous-divisions telles:

- le message
- le destinataire
- l'oblitération

# Première proposition d'encodage

```
<carte n="0010">
<recto url="Sables-olonne_recto_v.jpg"/>
<verso url="Sables-olonne_verso_v.jpg">
<obliteration>
<date>26-07-1967</date>
<lieu>Les Sables d'Olonne, Vendée</lieu>
</obliteration>
<message>
<p>Bientôt sur le chemin du retour avons eu du beau temps, tout le
mois de juillet. Vous souhaite à tous la même chose. J'espère que vous
êtes tous en bonne santé – A chacun son tour, le bon temps et le
repos. Avec d'affectueuses pensées et bons baisers à tous. Marcelle
</p>
</message>
<destinataire>
Mme Veyssade Lucienne
31, rue du Président Kennedy
ALFORTVILLE
Val de Marne
</destinataire>
</verso>
</carte>
```

## Commentaire

- On ne s'est pas servi du vocabulaire TEI: donc on aura de mal à expliquer cet encodage à des non francophones, ou de partager nos travaux avec des autres.
- On n'a pas mis en évidence tout ce qui pourrait nous intéresser de point de vue patrimoniale, par exemple les ratures, les retours à ligne, les noms de personnes ou de lieux, les aspects linguistiques, les metadonnées ...
- On n'a pas structuré (par exemple) le destinataire, qui va nous embêter pour faire des recherches efficaces...
- Pour ajouter de telles choses on doit inventer encore de vocabulaire... ce qui gaspille un peu notre temps, vue que la TEI l'a déjà fait...



## Proposition TEI initiale

On traite chaque carte comme un `<text>`

On traite chaque division comme un `<div type="[fonction]">`,  
facultativement divisé en `<p>`

Les noms de lieux seront balisés avec `<placeName>`, les noms de personnes avec `<persName>`, et les dates avec `<date>`

En plus, on se servira de `<del>` (pour les ratures), `<add>` (pour les ajouts), `<reg>` (pour les regularisations) `<hi>` pour les mots ou phrases mises en evidence.

On va structurer les composants du destinataire en `<address>`, qui regroupe un `<persName>` et des `<addrLine>`

On aura également besoin de l'attribut `@facs` pour associer une partie de texte avec son image numerise

Est-ce que cela nous servira?



# Premiere proposition d'encodage TEI

```
<text n="0010">
  <div type="recto" facs="Sables-olonne_recto_v.jpg"/>
  <div type="verso" facs="Sables-olonne_verso_v.jpg">
    <div type="obliteration">
      <date>26-07-1967</date>
      <placeName>Les Sables d'Olonne, Vendée</placeName>
    </div>
    <div type="message">
      <p>Bientôt sur le chemin du retour avons eu du beau temps, tout le
        mois de juillet. Vous souhaite à tous la même chose. J'espère que
vous
        êtes tous en bonne santé – A chacun son tour, le bon temps et le
        repos. Avec d'affectueuses pensées et bons baisers à tous.
Marcelle
      </p>
    </div>
    <div type="destinataire">
      <address>
        <persName>Mme Veyssade Lucienne </persName>
        <addrLine>31, rue du Président Kennedy </addrLine>
        <addrLine>ALFORTVILLE </addrLine>
        <addrLine>Val de Marne</addrLine>
      </address>
    </div>
  </div>
</text>
```

# Un ODD pour valider cela

```
<schemaSpec ident="tei_cartes">
  <moduleRef key="textstructure" include="TEI teiHeader text body div"/>
  <moduleRef
    key="core"
    include="date list item p address addrLine del add title reg lb hi
figure graphic"/>
  <moduleRef key="namesdates" include="persName placeName"/>
  <moduleRef
    key="header"
    include="fileDesc titleStmt publicationStmt sourceDesc revisionDesc"/>
  <moduleRef key="transcr" include="att.global.facs"/>
</schemaSpec>
```

Cette spécification de schéma est à compléter avec de la doc, bien sûr: voir fichier tei\_cartes.odd par exemple



## L'avantage ODD

Un ODD est un document TEI, qui peut être transformé en

- un schéma RNC, XSD, ou DTD pour valider les documents TEI
- documentation de ce schéma

Ces transformations s'effectuent avec le logiciel Roma, qui permet aussi

- l'exploration de tout l'univers TEI
- création et modification d'un ODD

On va expérimenter cela...

