



HAL
open science

Annoter et baliser

Emmanuelle Morlock

► **To cite this version:**

Emmanuelle Morlock. Annoter et baliser. École thématique. Ecole d'été du Consortium DISTAM 2024, L'arbresle, Couvent de la Tourette, France. 2024. hal-04774934

HAL Id: hal-04774934

<https://hal.science/hal-04774934v1>

Submitted on 9 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Annoter et baliser

Emmanuelle Morlock
IR CNRS – HiSoMA, Biblissima+

École d'été du consortium DISTAM - 9 juillet 2024 - Couvent de la Tourette

Annoter

Scholarly Primitives (J. Unsworth)

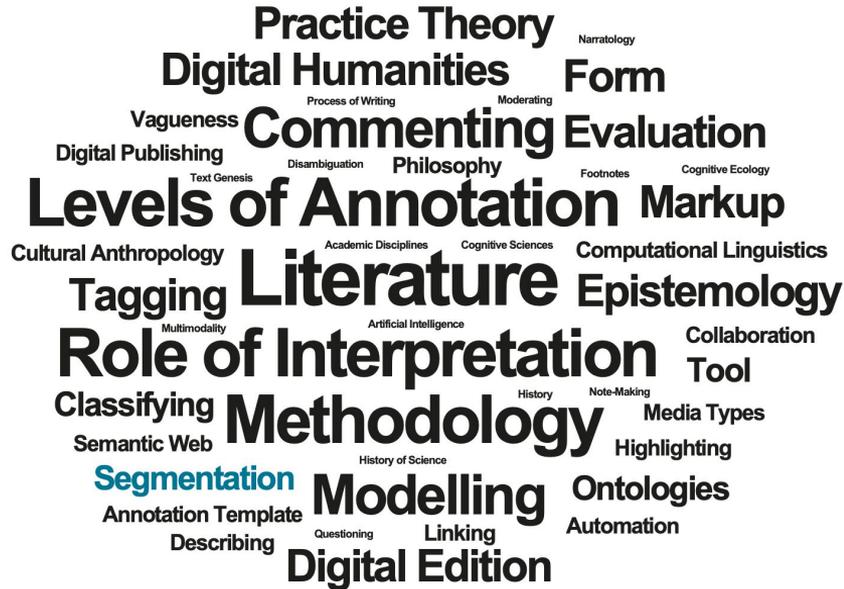
Les opérations fondamentales du travail savant :

- Découvrir
- **Annoter**
- Comparer
- Faire référence
- Prendre un échantillon
- Illustrer
- Représenter

Unsworth 2000 Unsworth, J. *Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?*, in *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. King's College, London (2000).

<https://johnunsworth.name/Kings.5-00/primitives.html>.

Les multiples facettes de l'annotation (concept théorique et appliqué)



Nantke and Schlupkothen 2020 Nantke, J. and F. Schlupkothen. Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization. De Gruyter, Berlin/New York (2020).

<https://doi.org/10.1515/9783110689112>.

Multiplicité des notions & questions posées par l'annotation

- Notes marginales manuscrites
- Notes de bas de page
- Surlignements / soulignements / coups de stabilo
- Commentaires vs notes explicatives
- Révisions
- Références à d'autres documents
- Catégories analytiques
- Indexation
- Distinction de niveaux narratifs dans les textes littéraires
- Multimodale (image, son, video)
- Comment les interpréter ? avec quels savoirs (auteur, contexte)
- Manuel / assisté par ordinateur : l'outil influence-t-il la pensée ?
- Privée ou publique (hypothes.is)
- Cercle herméneutique

Formule

A ajoute l'information *B*

à l'objet *C*

selon la modalité *D*

avec la finalité *E*.

L'annotation numérique

Lorsqu'ils interviennent en superposition d'un support préexistant, les actes d'annotation (écrire-tracer-dessiner) constituent le "**marquage**" instrumenté d'un texte et éventuellement de ses illustrations :

- Souligner
- Surligner
- Entourer
- Barrer
- Cocher
- Ecrire autour/à côté/en marge
- Schématiser une idée, un propos

<https://educavox.fr/formation/analyse/annoter-avec-le-numerique> (voir notamment la bibliographie)

Le balisage : pas nécessairement numérique

Mot à centrer

SALAGON

Lignes à centrer

*Conservatoire du patrimoine ethnologique
de Haute-Provence*

Lettre à changer

Le prieuré b**é**médic**n** de Salagon conserve
du XII^e siècle son église à deux vaisseaux. Un
logis ordonnancé de fenêtres à meneaux lui
a été **aj**olé au XV^e siècle.

/n
/XII

Mettre en petite capitales

Lettre à ajouter

Deux **co**urs caladées, des dépendances à
usage **f**ermier (XVI^e- XIX^e siècle), aujourd'hui
aménagées en **d'**exposition, complètent cet
ensemble classé **cl**assé monument histo-
rique.

lc
yo
→ agricole
↳ salles
→ y

Lettre à supprimer

Mot à changer

Mot à ajouter

Mot à supprimer

Baliser comme forme particulière d'annotation

Pourquoi a-t-on besoin d'un balisage spécial dans le numérique ?

Objectif :

Disposer d'un langage suffisamment expressif et précis pour nous permettre d'exprimer les données fondamentales d'un texte, de manière à ce que l'essentiel du sens véhiculé par ce texte puisse être transféré d'une application informatique à une autre...

Pourquoi nous avons besoin du balisage

- Fin des années 60 : idée de séparer le contenu (le texte) de la mise en forme
- Idée de développer de systèmes génériques pour favoriser l'échange des corpus / jeux de données
- 1986 : SGML devient norme ISO
- 1998 : XML recommandation du W3C
- 2007 : TEI P5 (abandon de SGML pour XML)

```
<!DOCTYPE document SYSTEM
"document.dtd">
  <DOCUMENT>
    <head>
      <title>Le titre</title>
      (...)
```

Structure d'une balise XML - 1

balise ouvrante



Contenu



balise fermante



<balise> chaîne de caractères **</balise>**

Structure d'un balise XML - 2

Attribut



`<balise type="simple">contenu</balise>`



valeur d'attribut

Les langages de balisage

- il en existe un très grand nombre
- A ne pas confondre avec les *langages de programmation* : les langages de balisage se contentent de *décrire* (ne donnent pas d'instructions)
- Ils explicitent des contenus par le biais :

- d'éléments
- d'attributs
- de texte

```
<book>  
  <title type = "last">Phèdre</title>  
  <title type = "first">Phèdre et Hyppolyte</title>  
  <author>Racine</author>  
    Le jour n'est pas plus pur que le fond de mon  
    coeur.  
</book>
```

Formalisation de la structure logique du texte



L'ÉCOLE DES FEMMES

ACTE PREMIER

SCÈNE PREMIÈRE CHRYSALDE, ARNOLPHE.

CHRYSALDE.

Vous venez, dites-vous, pour lui donner la main?

ARNOLPHE.

Oui, je veux terminer la chose dans demain.

CHRYSALDE.

Nous sommes ici seuls, et l'on peut, ce me semble,
Sans craindre d'être ouïs, y discourir ensemble.

Voulez-vous qu'en ami je vous ouvre mon cœur?

Votre dessein pour vous me fait trembler de peur,

2

Source gallica.bnf.fr / Bibliothèque nationale de France

```
<text>
```

```
<body>
```

```
<head>L'École des femmes <lb/><hi rend="1">Comédie</hi></head>
```

```
<div1 type="act" xml:id="1">
```

```
<head>Acte Premier</head>
```

```
<div2 type="scene" xml:id="I01">
```

```
<head>Scène Première</head>
```

```
<sp who="chrysalde" xml:id="I01-1">
```

```
<i>Vous venez, dites-vous, pour lui donner la main ?</i>
```

```
</sp>
```

```
<sp who="arnolphe" xml:id="I01-2">
```

```
<i>Oui, je veux terminer la chose dans demain.</i>
```

```
</sp>
```

```
<sp who="chrysalde" xml:id="I01-3">
```

```
<i>Nous sommes ici seuls, et l'on peut, ce me semble,</i>
```

```
<i>Sans craindre d'être ouïs, y discourir ensemble.</i>
```

```
<i n="5">Voulez-vous qu'en Ami je vous ouvre mon cœur ?</i>
```

```
<i>Votre dessein, pour vous, me fait trembler de peur :</i>
```

et aussi pour les humains...

<lb ed="T"/>何必能稱美談。」

<persName key="A001583">什</persName>每有疑義，必
共諮決。時偽 <lb ed="T"/>秦主 <persName
key="A002097">姚興</persName>專志經法，供養三千餘
僧， <placeName key="PL000000042002">關中
</placeName>舊僧 (...)

Comparez

- J'ai bien aimé *Mamma Mia*
- Elle a alors crié *Mamma Mia !*

- J'ai bien aimé **[titre]**Mamma Mia**[fin de titre]**
- Elle a alors crié **[autre langue]**Mamma Mia !**[fin autre langue]**

- Elle a alors crié **[autre langue : italien]**Mamma Mia !**[fin autre langue]**

Types de langages de balisages (*Markup Languages*)

Langage de balisage : système d'encodage de la structure, présentation ou sémantique des éléments composant un texte et de leurs relations.

Procédural	RED INK ON; PRINT "-1000 €"; RED INK OFF
Présentationnel	<code>\textcolor{red}{-1000€}</code>
Descriptif	<p>Vous êtes en découvert de :</p> <pre><measure unit="euro" value="-1000"> 1000 euros </measure></pre>

Rôle du balisage

Le balisage explicite aux machines ce qui est implicite pour les humains

- **Structures logiques et physiques des documents**
 - Livre > chapitre > paragraphe
 - Pièce de théâtre > acte > scène
- **Éléments indexables (entités nommées)**
 - noms, lieux, dates, etc.
- **Catégories d'analyse**
 - linguistiques
 - narratologiques,
 - interprétations spécifiques à une perspective de recherche

Balisage sémantique et non présentationnel

Le balisage présentationnel enregistre des apparences

- `<i>mens sana in corpore sano</i>`
- `<i>Le roman de la Momie</i>`
- C'est `<i>juste</i>` une impression...

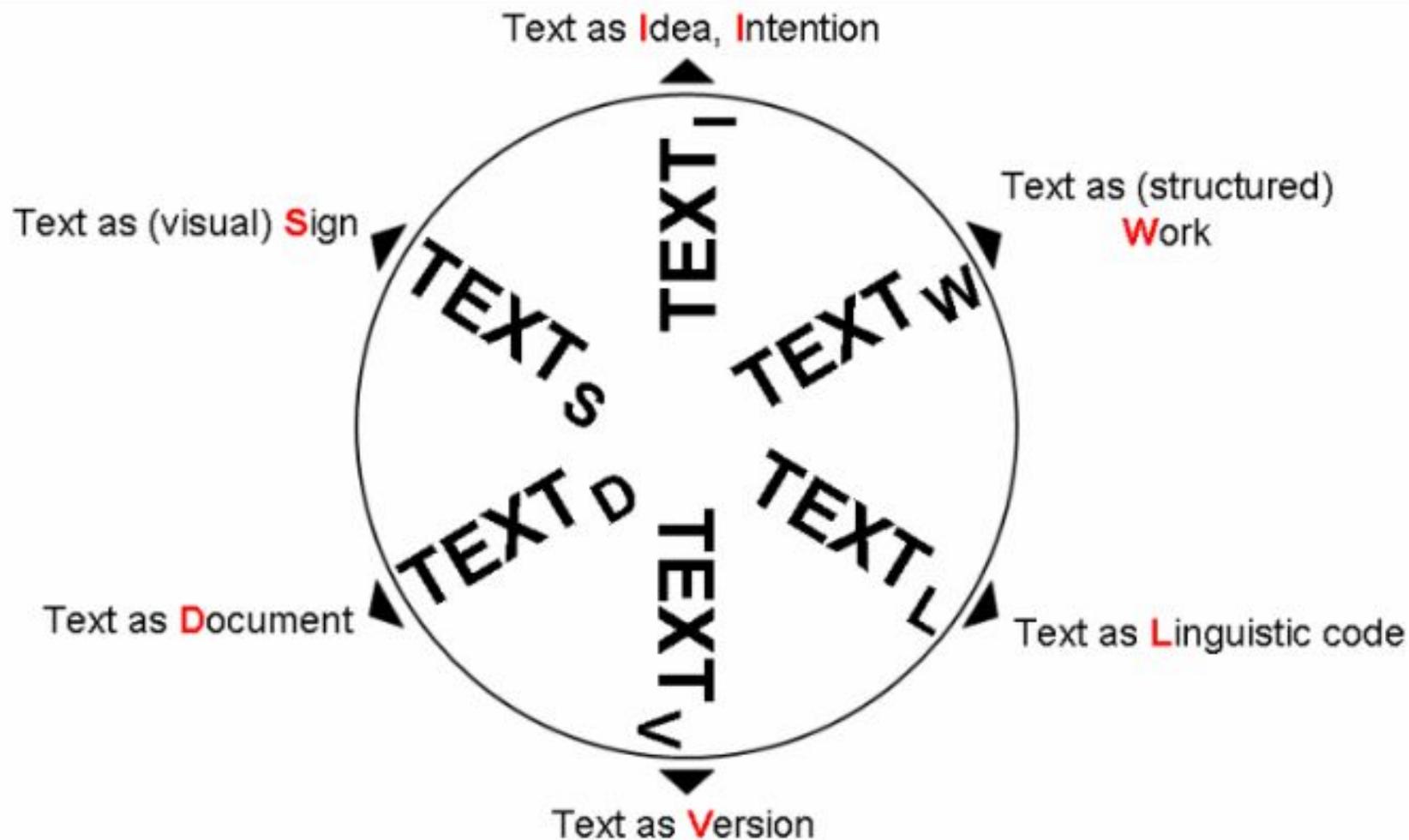
Le balisage sémantique capture le sens (ou la fonction)

- `<foreign>mens sana in corpore sano</foreign>`
- `<title>Le roman de la momie</title>`
- C'est `<emph>juste</emph>` une impression...

Un texte est une abstraction

- un document est un objet réel, qui a une existence dans le mode physique
- un texte est une abstraction, construite en référence à une communauté (ou une pratique sociale attestée)

- le texte =
le texte
 - + le point de vue sur le texte
 - + la perception de ce point de vue par un individu



Le balisage de texte, c'est donc :

- une étape essentielle pour créer une édition savante numérique
- la formalisation d'une interprétation (le texte + un point de vue sur le texte)
- produire une représentation lisible par les humains et par les machines
- expliciter les traits caractéristiques du document de manière à les rendre exploitables par des logiciels informatiques
- produire une source à partir de laquelle on peut générer autant de versions que l'on veut
 - dans un formatage adapté à différents appareils de lecture
 - dans des versions plus ou moins exhaustives ou transformées en fonction d'objectifs éditoriaux (version reproduisant l'orthographe du scripteur vs version normalisée)
- faciliter/permouvoir l'échange et la préservation des données (l'explicitation des implicites étant une condition de l'intelligibilité future de la représentation)

Les fonctions du balisage

- **Fonction structurelle**
 - explicite la structure
 - délimite les composants du document
- **Fonction descriptive**
 - donnent des informations
 - facilitent la recherche d'information
- **Fonction présentationnelle**
 - mettent en forme l'information dans un format visuel
 - ont pour but de favoriser la lisibilité
- **Fonction relationnelle**
 - liens hypertexte
 - lancement d'applications

exemple

Paris, le 13 septembre 2010

Cher Jeanne,
je vous remercie de la fabuleuse
soirée que j'ai passée en votre com-
pagnie à la Comédie française.
Les acteurs se sont surpassés
et j'ai particulièrement apprécié
cette mise en scène de Phèdre
Ce fut délicieux.

Henri

Balilage structurel

<ENTETE>Paris, le 13 septembre 2010**</ENTETE>**

<SALUTATIONS>Chère Jeanne,**<SALUTATIONS>**

<CORPS>

Je vous remercie pour la fabuleuse soirée que j'ai passée en votre compagnie à la Comédie Française. Les acteurs se sont surpassés et j'ai particulièrement apprécié cette mise en scène de Phèdre. Ce fut délicieux.

</CORPS>

<SIGNATURE>Henri**</SIGNATURE>**

Balisage présentationnel

<ALIGN-RIGHT>Paris, le 13 septembre 2010**</ALIGN-RIGHT>**

<BOLD>Chère Jeanne,**<BOLD>**

Je vous remercie pour la fabuleuse soirée que j'ai passée en votre compagnie à la Comédie Française. Les acteurs se sont surpassés et j'ai particulièrement apprécié cette mise en scène de

<UNDERLINE>Phèdre**</UNDERLINE>**.

Ce fut **<BLINK>**délicieux.**</BLINK>**

<FONT-FACE="Times">Henri**</FONT-FACE>**

Balisage descriptif

<PLACE>Paris</PLACE>, le <DATE>13 septembre 2010</DATE>

Chère <PERSON>Jeanne,</PERSON>

Je vous remercie pour la fabuleuse soirée que j'ai passée en votre compagnie à la <PLACE>Comédie Française</PLACE>. Les acteurs se sont surpassés et j'ai particulièrement apprécié cette mise en scène de <PLAY>Phèdre</PLAY>.

Ce fut <EMOTION>délicieux</EMOTION>.

<PERSON>Henri</PERSON>

Balisage relationnel

Paris, le `<TODAY href="date" actuate="auto"/>`

Chère Jeanne,

Je vous remercie pour la fabuleuse soirée que j'ai passée en votre compagnie à la Comédie Française. Les acteurs se sont surpassés et

j'ai particulièrement apprécié cette mise en scène de

``Phèdre``.

Ce fut délicieux.

``

``Henri``

Balisage procédural

```
\section{Formats produits par \LaTeX}
```

```
\label{memoformats}
```

Les principaux sont les suivants:

```
\begin{itemize}
```

```
\item pdf \item html (avec tth ou tex4ht ou d'autres) \item ps (vieux) \item rtf
```

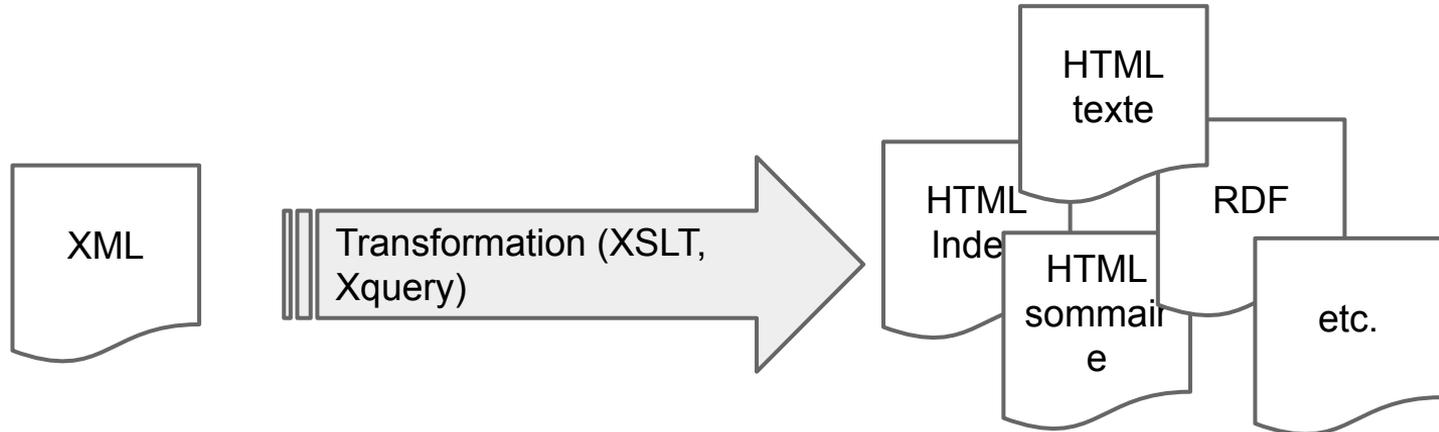
```
\end{itemize}
```

Fonctions du balisage comme “annotation structurée”

- délimiter (rendre “adressable”)
- expliciter (qualifier)
- représenter (une structure ou un phénomène, implicite ou explicite)
- ajouter de l’information (notes critiques par ex.)

Atouts du balisage descriptif

- Le “contenu” prime sur la mise en forme
- Cette distinction permet la ré-utilisation
- Elle est à la base de la *flexibilité* de l’édition numérique



Comparer

Mon amie `<persName>Adélaïde</persName>`

Je vis à `<placeName>Adélaïde</persName>` en Australie

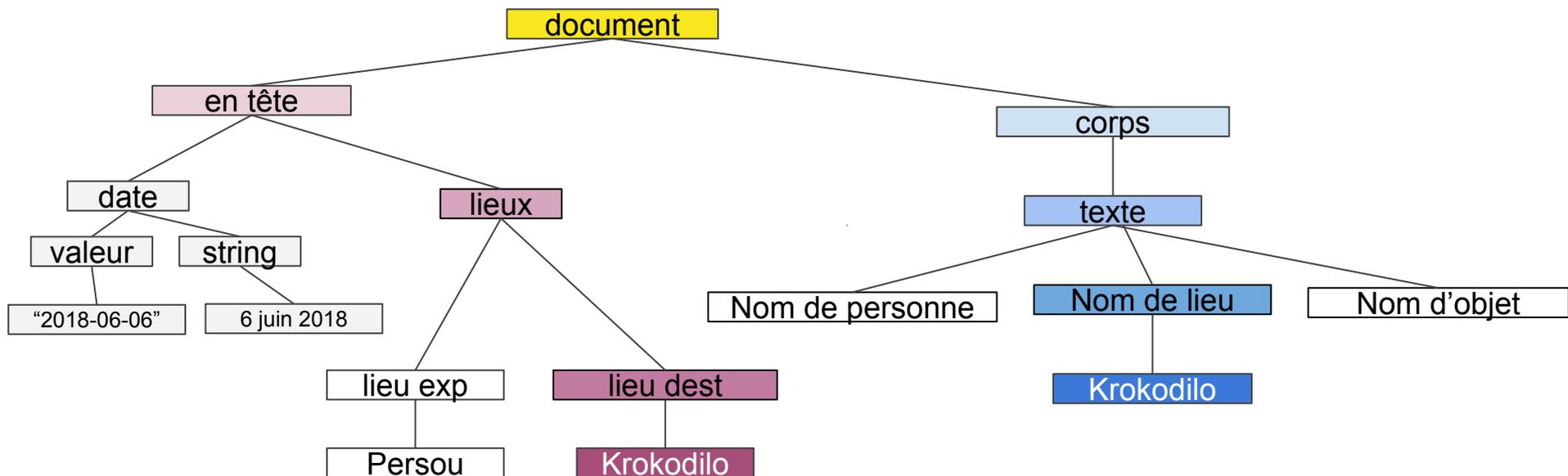
Que va-t-on baliser ?

- Première étape : analyse du document
- Choix des éléments que l'on souhaite représenter
 - aspects visuels (mise en page, forme des lettres, décorations...)
 - dimension linguistique et structurelle
 - information sur le monde réel (lieux, personnes, événements historiques...)
- Qu'est-ce qui est le plus important ?
 - *à vous de décider ! (a priori, selon la question de recherche)*

XML en 5 points

1. XML décrit les données mais ne les affiche pas
2. Les balises ne sont pas prédéfinies
3. Le balisage décrit la structure et le contenu
4. On doit utiliser des “grammaires” (schémas, DTD) pour contrôler le balisage
5. XML permet des traitements automatisés tout en restant lisible par des humains

Une hiérarchie ordonnée d'éléments de contenus



<recette>

<titre>Lait de poule</titre>

<ingredients>

<ingredient><qté>1</qté><nom>oeuf</nom></ingredient>

<ingredient><qté>10 cl</qté><nom>de lait chaud</nom></ingredient>

<ingredient><nom>sucre en poudre</nom></ingredient>

</ingredients>

<preparation>

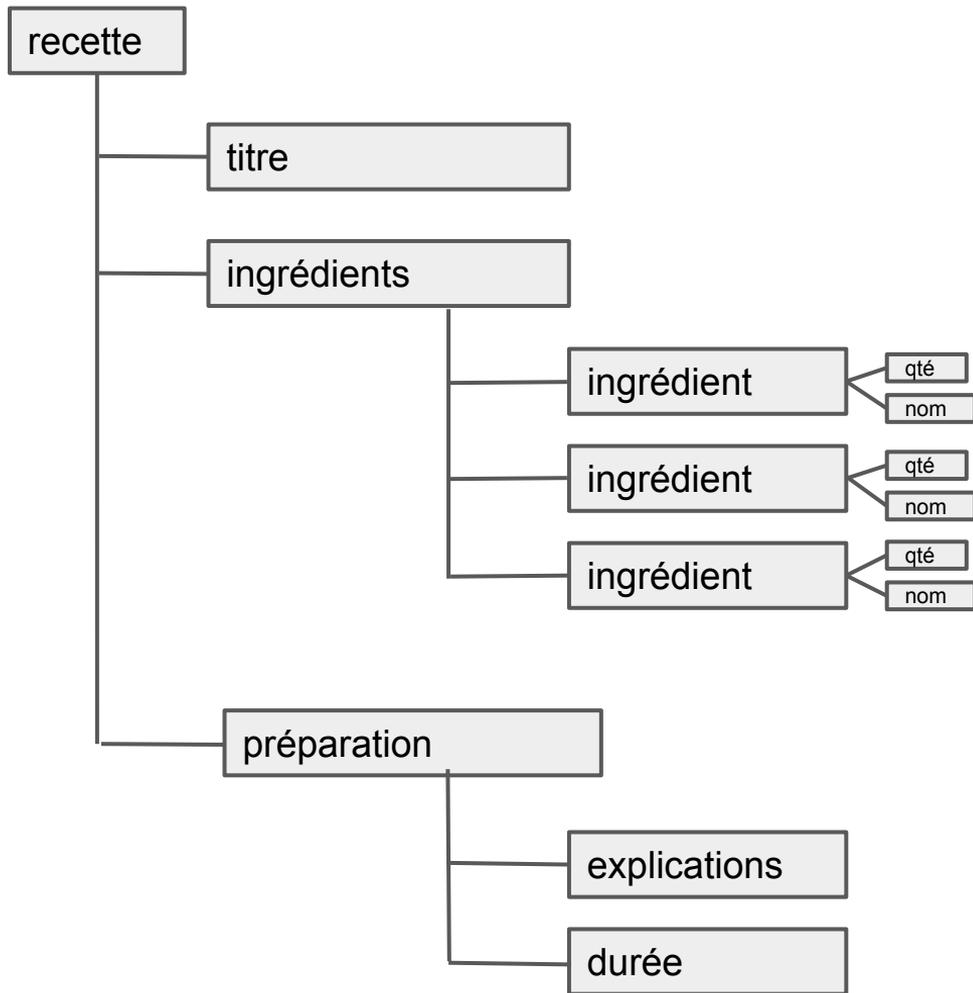
<explications>Verser le tout dans un verre à anse. Sucrez selon son goût. Remuer et ajouter un peu de noix de muscade râpée.

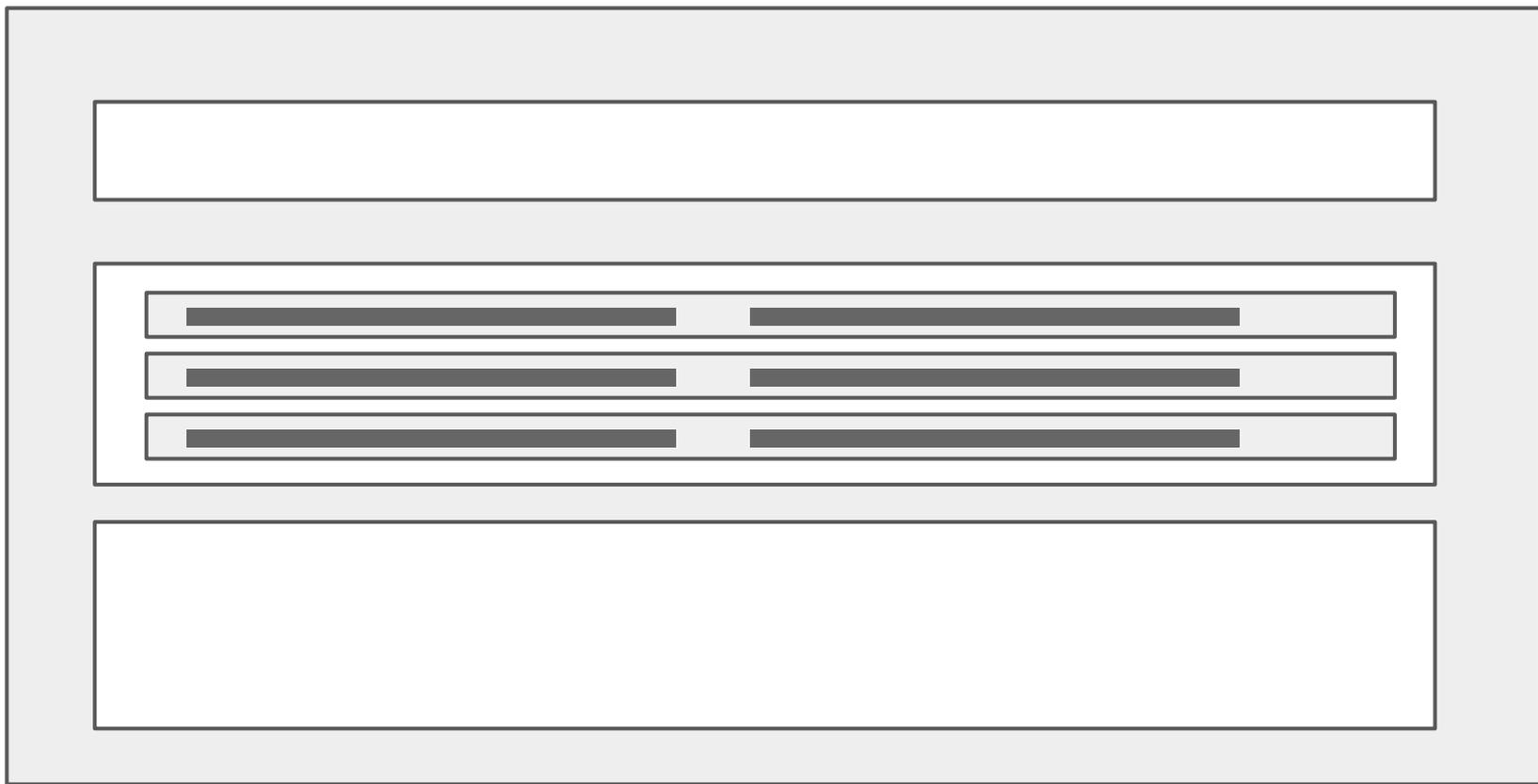
</explications>

<durée/>

</preparation>

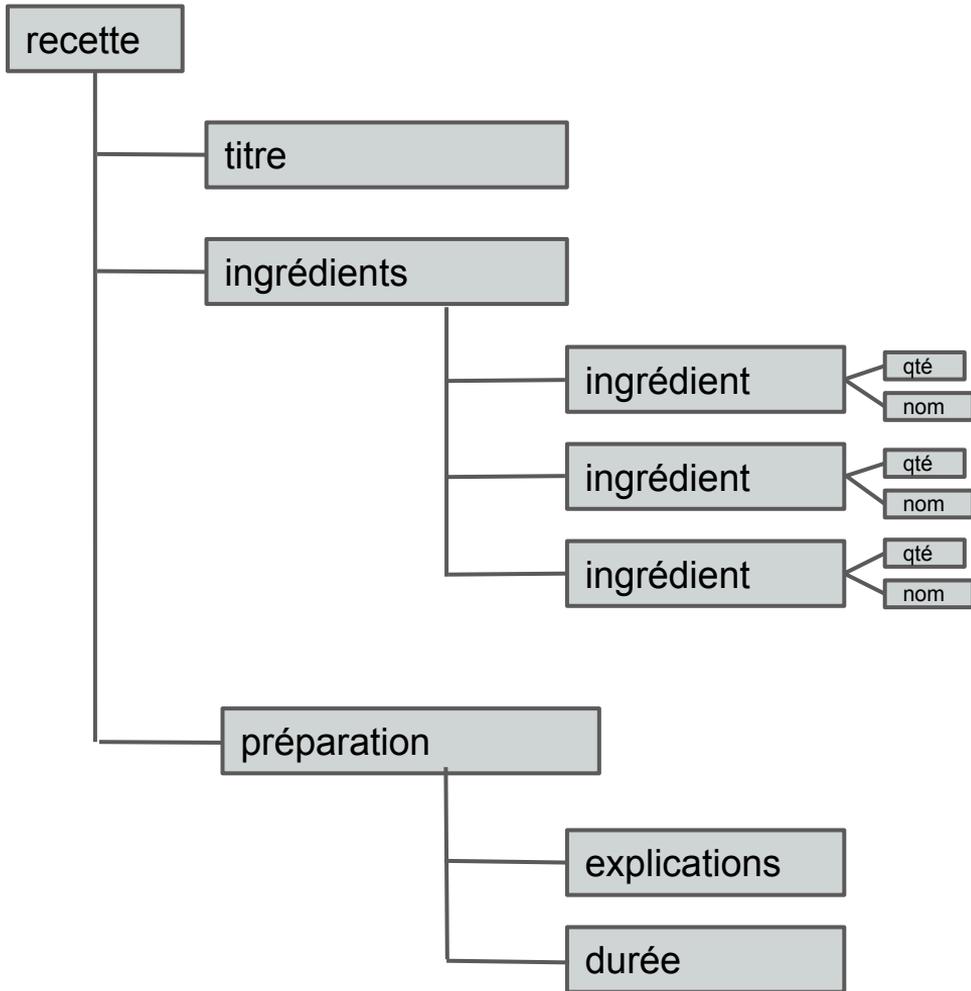
</recette>





[[lait de poule] [[1] [oeuf]]

[[10 cl] [de lait chaud]] [[Verser le tout dans un verre à
anse. Sucre selon son goût. Remuer et ajouter un peu de noix de muscade
râpée.]]



Conventions habituelles de l'épigraphie grecque et latine

...	Illegible letters, not restored by editor
աբգ	Letters ambiguous outside of their context
[...]	Letters missing, not restored by editor
[abc]	Letters missing, restored by editor
⟨...⟩	Letters erroneously omitted, not restored
⟨abc⟩	Letters erroneously omitted, not restored
a(bc)	Abbreviation expanded by editor
{abc}	Letters considered superfluous
[[abc]]	Deletion which can be restored

... et leur équivalent en TEI / EpiDoc

...	<code><gap reason="illegible" extent="unknown"/></code>
აბჟ	<code><unclear>abc</unclear></code>
[...]	<code><supplied reason="lost">abc</supplied></code>
[abc]	<code><supplied reason="lost">abc</supplied></code>
⟨...⟩	<code><gap reason="omitted" extent="unknown"/></code>
⟨abc⟩	<code><supplied reason="omitted">abc</supplied></code>
a(bc)	<code><expan><abbr>a</abbr><ex>bc</ex></expan></code>
{abc}	<code><surplus>abc</surplus></code>
[[abc]]	<code>abc</code>

Comment ?

Balisage “manuel”

- Editeur de texte basique (notepad, textedit)
- Editeur de texte outillé (oXygen, XMLMind, VScode...)
- Interfaces “auteur” (balisage par clic sur des boutons)

Balisage dans des chaînes de traitement

- Conversions (.docx stylé => TEI)
- Reconnaissance optique de caractères (HTR -> alto -> TEI)
- Autres outils (OpenRefine)

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://www.stoa.org/epidoc/schema/latest/tei-epidoc.rng"
  type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.stoa.org/epidoc/schema/latest/tei-epidoc.rng"
  type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="CK166" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmnt>
        <title>Statue assise de Ns-p3-mdw, fils de Ns-p3-q3-šwtj
          et de T3-ħ3'-n-B3stt, en scribe. Caire CG 48613</title>
      </titleStmnt>
    </fileDesc>
  </teiHeader>
</TEI>
```

Schema & documentation*

```
<body>
  <div type="text">
    <div type="edition">
      <div type="textpart">
        <ab>
          <supplied reason="lost"
            quantity="1" unit="character" />
          <supplied reason="lost"
            quantity="1" unit="character" extent="unknown" />
          <rdg source="#Boeckh">Αθηναίων</rdg>
          <rdg resp="#Kirchner">Αθην<supplied reason="lost">ναίων ἀναθείναι</supplied>
          <rdg source="#Erneher">Αθην<supplied reason="lost">ναίων καί</supplied>
```

- ▼ apparatus
- ▼ bibliography
- ▼ commentary
- ▼ edition
- ▼ textpart
- ▼ translation

to contain apparatus criticus or textual notes

eScriptorium Home Contact My Projects My Models Hello Ali

Manuscrits autographe de vingt mille lieues sous les mers / Jules Verne
 Document-1111.jpg - 270x500 - 131 kb

plus grand a coup sur que la rencontre d'un tigre en pleine forêt.

une demi heure après, guidés par la trainée électrique, nous atteignons le Nautilus. La porte extérieure était restée ouverte, et le capitaine Nemo la referma, des que nous fûmes rentrés dans la première cellule ; puis, il pressa un bouton. J'entendis manoeuvrer les pompes ; je sentis l'eau baisser autour de moi, et en quelques instants, la cellule fut entièrement vidée. La porte intérieure s'ouvrit alors, et nous passâmes dans le vestiaire.

Là, nos habits de scaphandre nous furent retirés, non sans peine, et, harassé, je regagnai ma chambre, encore tout émerveillé de cette excursion au fond des mers.

Chapitre 18.
Quatre mille lieues sous le Pacifique.

plus grand a coup sur que la rencontre d'un tigre en pleine forêt.

une demi heure après, guidés par la trainée électrique, nous atteignons le Nautilus. La porte extérieure était restée ouverte, et le capitaine Nemo la referma, des que nous fûmes rentrés dans la première cellule ; puis, il pressa un bouton. J'entendis manoeuvrer les pompes ; je sentis l'eau baisser autour de moi, et en quelques instants, la cellule fut entièrement vidée. La porte intérieure s'ouvrit alors, et nous passâmes dans le vestiaire.

Là, nos habits de scaphandre nous furent retirés, non sans peine, et, harassé, je regagnai ma chambre, encore tout émerveillé de cette excursion au fond des mers.

Chapitre 18.
Quatre mille lieues sous le Pacifique.

Alix Chagué, « eScriptorium : une application libre pour la transcription automatique des manuscrits », *Arabesques* [En ligne], 107 | 2022, mis en ligne le 13 octobre 2022, consulté le 09 juillet 2024. URL : <https://publications-prairial.fr/arabesques/index.php?id=3100>

Transformer le balisage pour la version de lecture

```
<choice>  
  <sic>relea</sic>  
  <corr cert="low">relatio</corr>  
  <corr cert="high">relicta</corr>  
</choice>
```

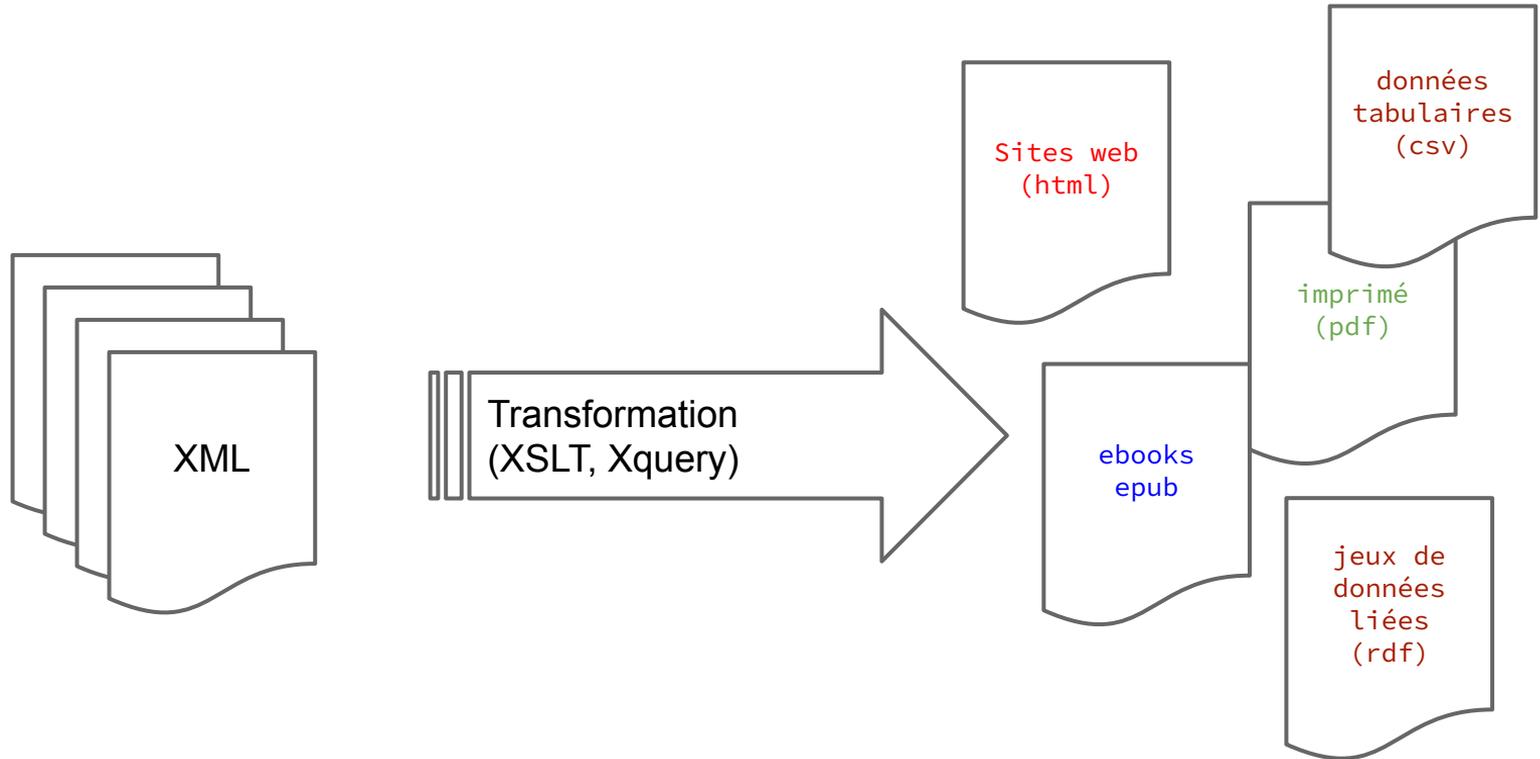
Transformation
(XSLT, Xquery)

relea

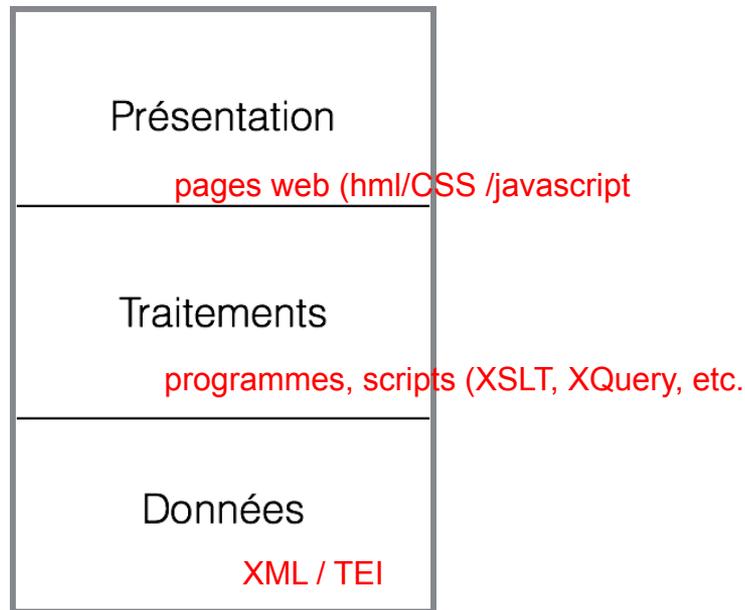
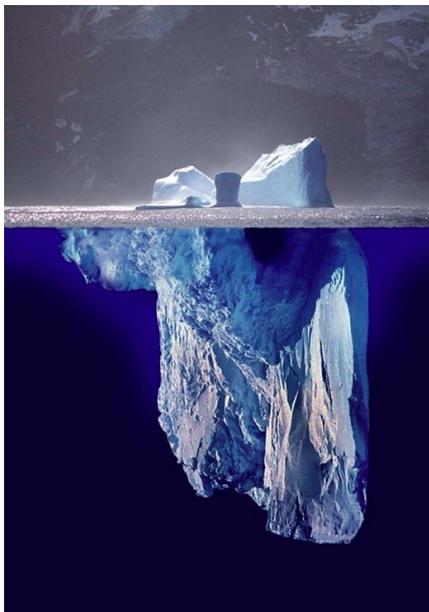
relatio

relicta

Edition multisupports

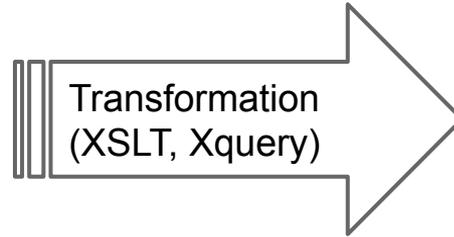


Les couches technologiques d'une édition en lignes



```
<xml>
<TEI>
(...)
<div>
<head>Atelier
pratique</head>
<p>L'encodage XML/TEI,
c'est facile !</p>
</div>

(...)
</TEI>
```



```
<html>
(...)
<h1>Atelier pratique</h1>
<p>L'encodage XML/TEI,
c'est facile !</p>

(...)

</html>
```

XML

Couche données

XSLT

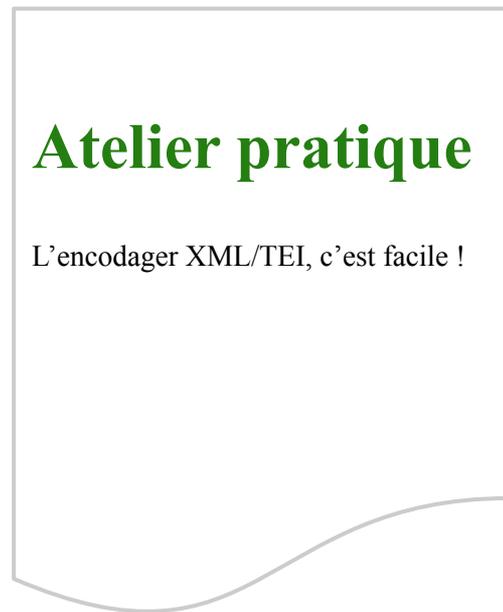
Couche "traitements"

HTML

Couche présentation

```
<html>
(...)
<h1>Atelier pratique</h1>
<p>L'encodage XML/TEI,
c'est facile !</p>
(...)
</html>
```

```
body {
  font-family:Times;
}
h1 {
  font-size: 200%;
  color: green;
  font-weight: bold;
}
p {
  color: black;
  font-size: 100%;
  margin-top:10%;
}
```



HTML
h1 = Titre de niveau 1



CSS
document : Police Times
titre 1 : taille 200%, vert
paragraphe : taille 100% noir



Affichage dans le navigateur web

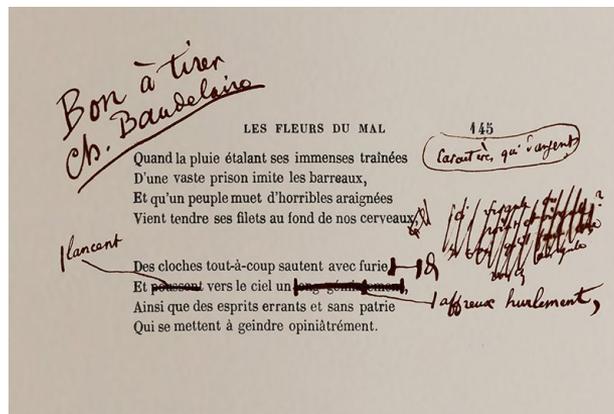
Que peut-on représenter ?

structure du document (titres, paragraphes, strophes)

Structures linguistiques (parties du discours, phrases, mots...)

Références à des personnes ou objets du monde réel, des événements

Métadonnées (auteur, date de création, provenance, genre)



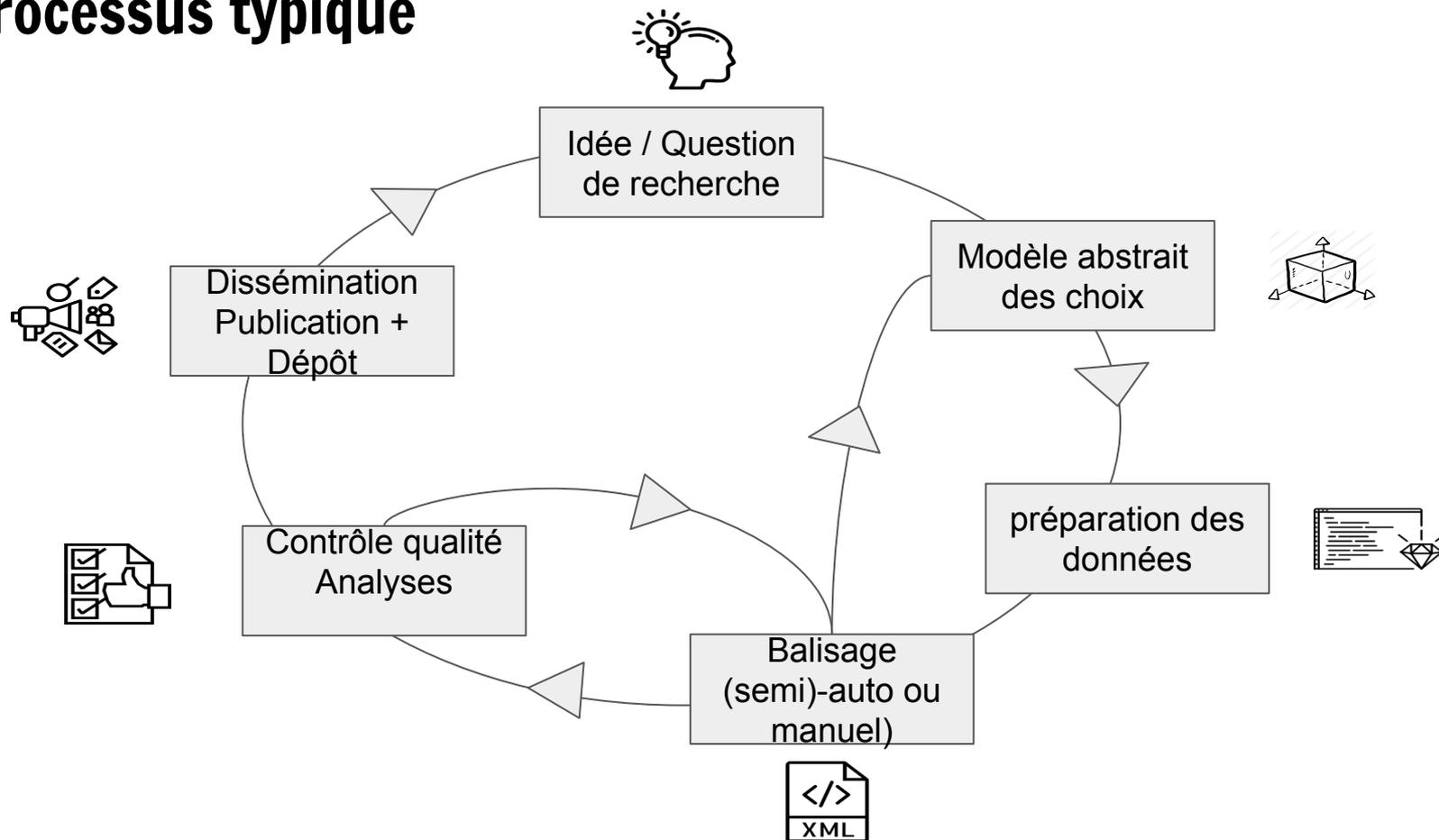
Caractéristiques du document support (dimensions, état du support, couleur, type de papier...)

Disposition, mise en page, analyse des formes graphiques et manuscrites

Processus : ratures, campagnes de correction

Annotations, notes de régie...

Processus typique



3 mécanismes clés du balisage XML/TEI

Mécanisme clé : identifier / pointer

```
<text>
  <body>
    <div n="1">
      <p n="1"> [...] Vade retro Satana, scandalum
<anchor xml:id="#app1b"/>es nobis <anchor
xml:id="#apple"/>
      </p> [...]
    </div>
  </body>
</text>
```

```
<app from="#app1b" to="#apple">
  <rdg wit="#B">est mihi</rdg>
</app>
```

Mécanisme : parallélisation

```
<text>
  <body>
    <div n="sermo1">
      <p n="1"> [...] Vade retro Satana, scandalum
        <app>
          <lem>es nobis</lem>
          <rdg wit="#B">est mihi</rdg>
        </app> [...]
      </p> </div>
    </body>
  </text>
```

Balisage “débarqué” (*standoff markup*)

- élément conteneur <standOff>
- pour :
 - des données liées
 - de l'information contextuelle
 - des listes d'annotations...

- utilise le mécanisme “identification / pointeur”

```

<text>
  <body>
<!-- ... -->
  <p>
    <w xml:id="w51">I</w>
    <w xml:id="w52">wanna</w>
    <w xml:id="w53">put</w>
    <w xml:id="w54">up</w>
    <w xml:id="w55">new</w>
    <w xml:id="w56">wallpaper</w>
  <pc>.</pc>
  </p>
<!-- ... -->
</body>
</text>
<standOff type="morphosyntax">
  <spanGrp type="wordForm">
    <span target="#w51" ana="#fs01"/>
    <span target="#w52" ana="#fs02"/>
    <span target="#w52" ana="#fs03"/>
    <span target="#w53 #w54" ana="#fs04"/>
    <span target="#w55" ana="#fs05"/>
    <span target="#w56" ana="#fs06"/>
  </spanGrp>
  <fs xml:id="fs01">
    <f name="lemma">
      <string>I</string>
    </f>
    <f name="pos">
      <symbol value="PP"/>
    </f>
  </fs>
  <fs xml:id="fs02">
    <f name="lemma">
      <string>want</string>
    </f>
    .. ..

```

```
<text>
  <body>
    <div type="edition">
      <div type="textpart" subtype="chapter"
        n="1" xml:id="ch1">
        <p n="1" xml:id="c1p1">
          <seg n="1" xml:id="c1p1s1">Gallia est omnis divisa in partes tres, quarum unam
            incolunt Belgae, aliam Aquitani, tertiam qui ipsorum lingua Celtae, nostra
            Galli appellantur.</seg>
        <!-- ... -->
          <seg n="6" xml:id="c1p1s6">Belgae ab extremis Galliae finibus oriuntur,
            pertinent ad inferiorem partem fluminis Rheni, spectant in septentrionem et
            orientem solem.</seg>
          <seg n="7" xml:id="c1p1s7">Aquitania a Garumna flumine ad Pyrenaeos montes et
            eam partem Oceani quae est ad Hispaniam pertinet; spectat inter occasum solis
            et septentriones.]</seg>
        <!-- ... -->
          </p>
        <!-- ... -->
      </div>
    <!-- Ch. 2 etc. ... -->
  </div>
</body>
</text>
```

```
<standOff>
  <listAnnotation>
    <annotation xml:id="ann01"
      motivation="linking" target="#Gallia">
      <respStmt xml:id="ed">
        <resp>creator</resp>
        <persName>Fred Editor</persName>
      </respStmt>
      <revisionDesc>
        <change status="created"
          when="2020-05-21T13:59:00Z" who="#ed"/>
        <change status="modified"
          when="2020-05-21T19:48:00Z" who="#ed"/>
      </revisionDesc>
      <licence target="http://creativecommons.org/licenses/by/3.0/">
<!-- Gallia in seg 1 -->
        <ptr target="#string-range(c1p1s1,0,6)"/>
<!-- Galliae in seg 6 -->
        <ptr target="#string-range(c1p1s6,19,7)"/>
      </annotation>
<!-- ... -->
    </listAnnotation>
  </standOff>
```

Le balisage peut être (semi-)automatisé...

Inventory no. (EAD)	W	D	H	W	Langue(s)	Original locality	Present locality	Primary bibliography	Secondary bibliography	Visual documentation	Substance of inscription	Remarks	
1	31					Nagarjunakonda Site 124 (?)	NM 79	Sarkar 1963-66.	Sarkar 1960: 84 (postscript by editor). — AN 1961-62: 77. — AIE 1962-63: B.147. — Raghunath 2001: 67-68 (1).	Sarkar, Raghunath. — photos, JM 2017.	gift of a pillar	ARE states: "Slab discovered in Site No. 125, near the Royal Burning Chhat No. 127"	
2	20					Rensali, Palnad Taluk, Guntur District	unknown	Sankaranarayanan 1967-68.	AIE 1936-37: 62, B.149. — Raghunath 2001: 69 (2).	Est. ASI: Sankaranarayanan; Srinivasan & Sankaranarayanan (p).	erection of a stone Ayaka pillar	Village called Rentsaliracata nowadays.	
3	22	unavailable	unavailable	50	MA	Kintalapudi, Palnad Taluk, Guntur District	unknown	Sankaranarayanan 1969. — Khan Sankaranarayanan Rao et al. 1998: 178 (no. 1).	Raghunath 2001: 70 (3). — Hanumantha Rao et al. 1998: 178 (no. 1).	Subrahmanyam, Khan, p. 200; Sankaranarayanan; Raghunath. — photos	erection of a Buddha statue	slab written lengthwise, parallel to the boundaries between the facets of the stone. Cf. 036, and inscription at Nat. Mus. Colombo.	
4	47	38	44	55	MA	Nagarjunakonda Site 1, West of the Mahastupa	NM 301, 502	Vogel 1929-30a: 1 (C2). — Burgess 1963: 119-21.	Raghunath 2001: 72 (4), 194 (83).	estampage Leden 87. — Vogel.	gift for welfare and happiness	We have only identified the upper fragment; measurements concern only the	
5	46	30	88	56	MA	Nagarjunakonda Site 1, West side of the Mahastupa	NM 301, 502	Vogel 1929-30a: 1 (C2). — Burgess 1963: 119-21.	Raghunath 2001: 74 (5).	estampage Leden N6 (bottom), NF (top). — Vogel; Raghunath. — photos, JM 2017.	erection of a stone pillar	The estampages shown by Raghunath under his 5 and 6 together belong to one BADD03.	
6	48	29	105	55	MA	Nagarjunakonda Site 1, West side of the Mahastupa	NM 283, 299	Vogel 1929-30a: 1 (C2). — Burgess 1963: 119-21.	estampage Leden N6. — Vogel.	erection of a pillar for welfare and happiness		No rubbing for our BADD06 in Raghunath. See remark under BADD03.	
7	45	39	45	54	MA	Nagarjunakonda Site 1, West side of the Mahastupa	west Ayaka platform, reconstructed Mahastupa at	Vogel 1929-30a: 1 (C5).	estampage Leden N6. — Vogel; Raghunath.	erection of a pillar for welfare and happiness			
8	47	36	66	51	MA	Nagarjunakonda Site 1, South side of the Mahastupa	NM 285	Vogel 1929-30a: 13, 18-9 (85).	Raghunath 2001: 80 (17).	estampage Leden N6. — Vogel; Raghunath.	erection of a stone pillar for welfare and happiness		Raghunath (2001: 107) indicates Site #, but this is error for Site 1.

Tableur

fichier balisé
EpiDoc

```

<physDesc>
  <objectDes>
    <supportDes>
      <support>
        <p>
          <objectType>Ayaka pillar</objectType> in fragmentary state;
          <dimensions, unit="cm">
            <height>61</height>
            <width>40</width>
            <depth>13</depth>
          </dimensions>. </p>
        </support>
      </supportDes>
    </objectDes>
  </physDesc>
  <layoutDesc>
    <layout writtenLines="7">7 lines. Inscribed area: <dimensions>
      <height>40</height>
      <width>49</width>
    </dimensions>. </layout>
  </layoutDesc>
</objectDes>
    
```

Metadata

Support Ayaka pillar; h. 136 × w. 41 × d. 20 cm.

Text Middle Indo-Aryan, Southern Brāhmi script, 6 lines. Inscribed area unknown. In EIAD 31, 32, 33 as in 90, the engraver used a form of <cha> that is uncommon. The distinction between <cha> and <ḥa> in these cases appears to be faint.

Date The 20th regnal year of Siri-Virapurisadatta, i.e., approximately between 225 and 275 CE.

Origin Originally erected on the eastern Ayaka platform of the Stūpa near Jaggayyapeta.

Provenance Discovered during excavations organised by Burgess in February 1882. Identified at the Chennai Government Museum on 7 August 2017.

Visual Documentation Photo(s): AO 2017
Photo(s) of estampage(s): Burgess 1887, plate LXII (1)

Editors Arlo Griffiths and Vincent Tournier, with contributions by Stefan Baums and Ingo Strauch.

Publication history First described and edited by Bühler 1882: 258-9 (I), followed by Burgess 1882: 55-6 (reading and translation by Bhagvanlal Indraji) and then Burgess 1887: 110-1 (1) (new reading, with Bühler's translation). Re-edited here from the estampage published by Burgess.

Interface de lecture

Prototype de visualisation de données partir d'un corpus d'ostraca

ΑΠΟΛΛΩΣ ΠΡΙΣΚΟΥ

παντὸς ὑγειαίνιν· ἐ

Χαιρήμωνι ἱμιμάτιν

ὀβολοί· κώμιζε πο

δέσμην κράμ'πι

πλίω γὰρ τὴν δέσμη

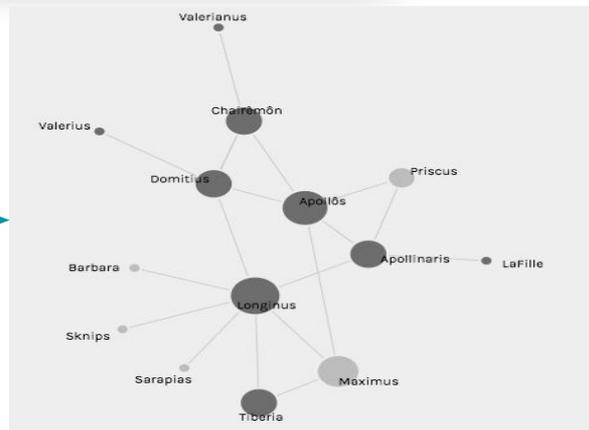
Μάξιμος τίρον καὶ Δομίτιος καὶ Ἀπολιναρίω·

ἀσπάζεστέ ἐς Ἀπολιναρίου τὸν φίλον σοῦ.

Apollôs à Priscus, son frère, salut et longue santé. J'ai reçu des mains de Chairêmôn un demi-mation de dattes et sept oboles. Reçois, via Capito le carrier, une botte de choux et un peu de rue. Il ne transporte pas plus que la botte. Salue Maximus la recrue, Domitius et Apollinaris. Apollinaris, ton ami, te salue. Porte-toi bien.

<persName>Apollôs</persName> à
 <persName>Priscus</persName>, son frère.
 <seg ana="#paiement">J'ai reçu</seg>
 des mains de <persName>Chairêmôn</persName>
 un <measure type="currency" unit="mation"><num
 value="0.5">demi</num>-mation</measure>
 de <seg ana="#objet-denrée">dattes</seg> et
 <measure type="currency" unit="obole">
 <num value="7">sept</num> oboles
 </measure>

	κ	χ	δ	δ	ε	σ	υ	ι	ε	λ	ν					
1	net	astax	Scipius	Domitius	Capitulus	his	dear	his	price	ex	exchange	clat	exchange	clat	exchange	intermediary
2	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Chairêmôn	Domitius	don	vas	karatu					
3	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Domitius	Chairêmôn	don	vas	karatu					
4	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Maximus	Chairêmôn	palmet	chutras	Maximus					
5	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Maximus	Chairêmôn	adrese	salut	Chairêmôn					
6	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Longinus	Domitius	adrese	salut	Chairêmôn					
7	O.Kro248	Apollis	Domitius	Chairêmôn	Priscus	Kivocid	Longinus	Apollinaris	adrese	salut	Chairêmôn					
8	O.Kro248	Apollis	Priscus	Apollis	Domitius	Kivocid	Chairêmôn	Apollis	adrese	salut	Chairêmôn					
9	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Apollis	Priscus	don							
10	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Apollis	Priscus	don							
11	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Apollis	Maximus	adrese	salut	Priscus					
12	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Apollis	Domitius	adrese	salut	Priscus					
13	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Apollinaris	Priscus	adrese	salut	Apollis					
14	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Longinus	Apollinaris	adrese	salut	Priscus					
15	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Longinus	Apollinaris	adrese	salut	Priscus					
16	O.Kro242	Apollis	Priscus	Apollis	Priscus	Kivocid	Longinus	Apollinaris	adrese	salut	Priscus					
17	O.Kro248	Longinus	Apollinaris	Longinus	Priscus	Kivocid	Longinus	Apollinaris	price	karat	Longinus					
18	O.Kro248	Longinus	Apollinaris	Longinus	Priscus	Kivocid	Longinus	Maximus	adrese	salut	Apollinaris					



Balisage

Le balisage permet deux choses fondamentales :

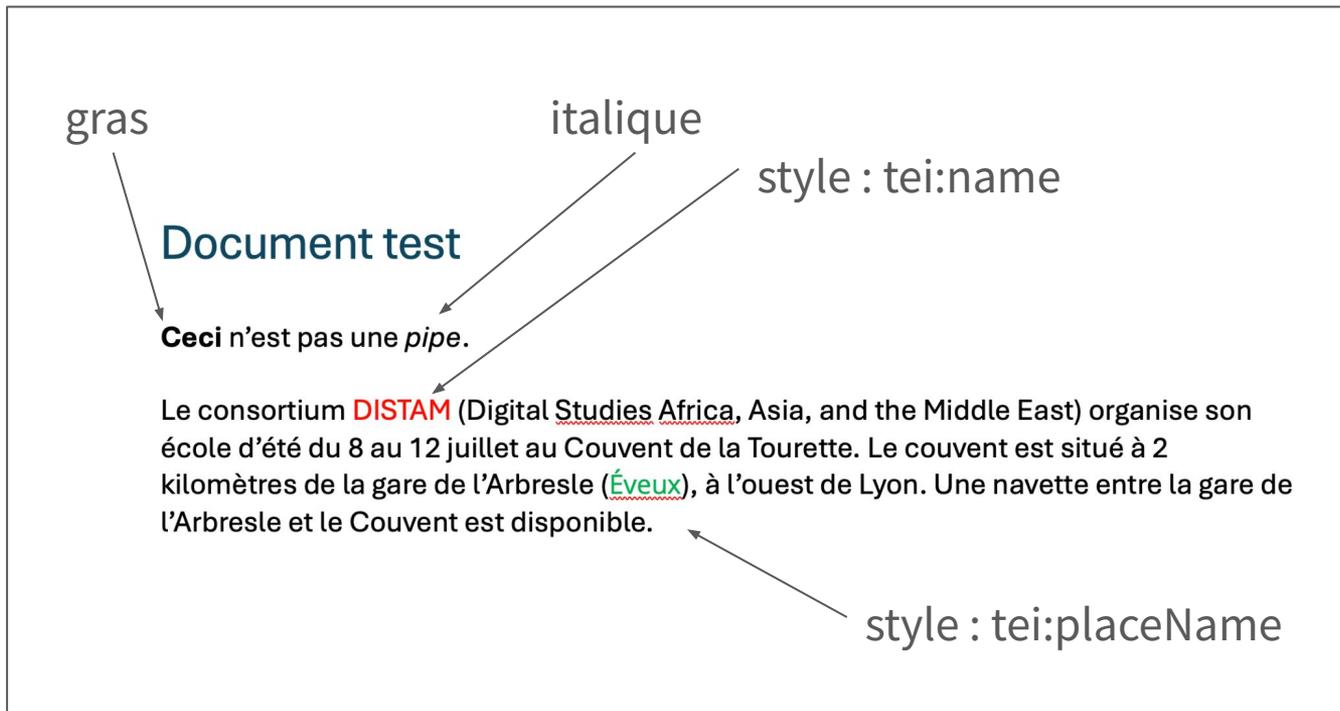
- Délimiter des unités “discrètes” au sein du texte
- Expliciter de manière univoque l’ajout d’information

Le langage de balisage :

- Un vocabulaire de balises (<p>, <div>, <note>, <opener>, <unclear>...)
- Une “grammaire” qui spécifie les relations possibles
 - enfants possibles
 - attributs possibles
 - et autres contraintes

Démos

Document word stylé



Démos

- TEI Publisher :
 - url : <https://teipublisher.com/>
 - bac à sable : <https://teipublisher.com/exist/apps/tei-publisher/index.html>
- Import d'un document Word stylé
- Exploitation d'un encodage parallèle (édition critique) et visualisation des témoins via IIF : *Sermones Demo*
- Passer de données tabulaires à un balisage TEI via l'interface graphique d'OpenRefine

```
        "width": 1327,
        "height": 2143
      },
      "on": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/canvas/p0005"
    }
  ],
  "thumbnail": {
    "@id": "https://digi.vatlib.it/pub/digit/MSS\_Reg.lat.245/thumb/Reg.lat.245\_0005\_fa\_0001r.tif.jpg",
    "@type": "dctypes:Image",
    "format": "image/jpeg",
    "width": 97,
    "height": 150
  },
  "viewingHint": "non-paged"
},
{
  "@id": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/canvas/p0006",
  "@type": "sc:Canvas",
  "label": "1v",
  "width": 1297,
  "height": 2141,
  "images": [
    {
      "@id": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/res/anno0006",
      "@type": "oa:Annotation",
      "motivation": "sc:painting",
      "resource": {
        "@id": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/res/img0006",
        "@type": "dctypes:Image",
        "format": "image/jpeg",
        "service": {
          "@context": "http://iiif.io/api/image/2/context.json",
          "@id": "https://digi.vatlib.it/iiifimage/MSS\_Reg.lat.245/Reg.lat.245\_0006\_fa\_0001v.jp2",
          "profile": "http://iiif.io/api/image/2/level2.json"
        },
        "width": 1297,
        "height": 2141
      },
      "on": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/canvas/p0006"
    }
  ],
  "thumbnail": {
    "@id": "https://digi.vatlib.it/pub/digit/MSS\_Reg.lat.245/thumb/Reg.lat.245\_0006\_fa\_0001v.tif.jpg",
    "@type": "dctypes:Image",
    "format": "image/jpeg",
    "width": 95,
    "height": 150
  }
},
{
  "@id": "https://digi.vatlib.it/iiif/MSS\_Reg.lat.245/canvas/p0007",
  "@type": "sc:Canvas",
  "label": "2r",
  "width": 1327,
  "height": 2143,
  "images": [
```


OpenRefine manifest json [Permalien](#)

Ouvrir... Exporter ▾ Aide

 > **294 lignes**

 Extensions **VIB-Bits** ▾ **RDF Transform** ▾ **RDF** ▾ **Wikidata** ▾

 Voir en: **lignes** entrées Afficher: **5** 10 25 50 100 500 1000 lignes

 « première < précédente 1 suivante > dernière »

<input type="checkbox"/> Toutes	<input type="checkbox"/> @id	<input type="checkbox"/> url	<input type="checkbox"/> id	<input type="checkbox"/> n	<input type="checkbox"/> type
 	1.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0001_al_piatto.anteriore.jp2	MSS_Reg.lat.245/Reg.lat.245_0001_al_piatto.anteriore.jp2	al_piatto.anteriore	
 	2.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0002_ba_risguardia.anteriore.jp2	MSS_Reg.lat.245/Reg.lat.245_0002_ba_risguardia.anteriore.jp2	ba_risguardia.anteriore	
 	3.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0003_cy_0001r.jp2	MSS_Reg.lat.245/Reg.lat.245_0003_cy_0001r.jp2	cy_0001r	1r
 	4.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0004_cy_0001v.jp2	MSS_Reg.lat.245/Reg.lat.245_0004_cy_0001v.jp2	cy_0001v	1v
 	5.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0005_fa_0001r.jp2	MSS_Reg.lat.245/Reg.lat.245_0005_fa_0001r.jp2	fa_0001r	1r recto
 	6.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0006_fa_0001v.jp2	MSS_Reg.lat.245/Reg.lat.245_0006_fa_0001v.jp2	fa_0001v	1v verso
 	7.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0007_fa_0002r.jp2	MSS_Reg.lat.245/Reg.lat.245_0007_fa_0002r.jp2	fa_0002r	2r recto
 	8.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0008_fa_0002v.jp2	MSS_Reg.lat.245/Reg.lat.245_0008_fa_0002v.jp2	fa_0002v	2v verso
 	9.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0009_fa_0003r.jp2	MSS_Reg.lat.245/Reg.lat.245_0009_fa_0003r.jp2	fa_0003r	3r recto
 	10.	https://digi.vatlib.it/iiifimage/MSS_Reg.lat.245/Reg.lat.245_0010_fa_0003v.jp2	MSS_Reg.lat.245/Reg.lat.245_0010_fa_0003v.jp2	fa_0003v	3v verso

Export du modèle

Préfixe

```
<facsimile>  
  <surfaceGrp>
```

Modèle de ligne

```
<surface type="{type}" n="{n}" xml:id="{id}">  
  <graphic url="{url}"/>  
</surface>
```

Séparateur de ligne

Suffixe

```
  </surfaceGrp>  
</facsimile>
```

```
<facsimile>  
  <surfaceGrp><surface type="null" n="" xml:id="al_piatto.anteri  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0001_al_piatto.anteri  
  </surface><surface type="null" n="" xml:id="ba_risguardia.anterior  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0002_ba_risguardia.a  
  </surface><surface type="null" n="1r" xml:id="cy_0001r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0003_cy_0001r.jp2"/>  
  </surface><surface type="null" n="1v" xml:id="cy_0001v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0004_cy_0001v.jp2"/>  
  </surface><surface type="recto" n="1r" xml:id="fa_0001r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0005_fa_0001r.jp2"/>  
  </surface><surface type="verso" n="1v" xml:id="fa_0001v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0006_fa_0001v.jp2"/>  
  </surface><surface type="recto" n="2r" xml:id="fa_0002r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0007_fa_0002r.jp2"/>  
  </surface><surface type="verso" n="2v" xml:id="fa_0002v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0008_fa_0002v.jp2"/>  
  </surface><surface type="recto" n="3r" xml:id="fa_0003r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0009_fa_0003r.jp2"/>  
  </surface><surface type="verso" n="3v" xml:id="fa_0003v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0010_fa_0003v.jp2"/>  
  </surface><surface type="recto" n="4r" xml:id="fa_0004r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0011_fa_0004r.jp2"/>  
  </surface><surface type="verso" n="4v" xml:id="fa_0004v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0012_fa_0004v.jp2"/>  
  </surface><surface type="recto" n="5r" xml:id="fa_0005r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0013_fa_0005r.jp2"/>  
  </surface><surface type="verso" n="5v" xml:id="fa_0005v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0014_fa_0005v.jp2"/>  
  </surface><surface type="recto" n="6r" xml:id="fa_0006r">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0015_fa_0006r.jp2"/>  
  </surface><surface type="verso" n="6v" xml:id="fa_0006v">  
    <graphic url="MSS_Reg.lat.245/Reg.lat.245_0016_fa_0006v.jp2"/>  
  </surface><surface type="recto" n="7r" xml:id="fa_0007r">
```

Réinitialiser le modèle

Exporter

Annuler

```
<facsimile>
  <surfaceGrp>

    <surface type="null" n="" xml:id="al_piatto.anteriore">
      <graphic url="MSS_Reg.lat.245/Reg.lat.245_0001_al_piatto.anteriore.jp2"/>
    </surface>

    <surface type="verso" n="10v" xml:id="fa_0010v">
      <graphic url="MSS_Reg.lat.245/Reg.lat.245_0024_fa_0010v.jp2"/>
    </surface>

    <surface type="recto" n="11r" xml:id="fa_0011r">
      <graphic url="MSS_Reg.lat.245/Reg.lat.245_0025_fa_0011r.jp2"/>
    </surface>

    (...)

  </surfaceGrp>
</facsimile>
```

Conclude ?

Quelques remarques en guise de conclusion

- Baliser c'est annoter en fonction d'objectifs d'exploitations
- Balisage "à la main" : utile à pratiquer pour apprendre / comprendre
- Tout part de l'analyse du document et des objectifs d'exploitation, en fonction d'une question de recherche
- Le balisage sémantique manuel de sources anciennes : quand l'expertise humaine est (encore) nécessaire...
- Granularité du balisage / nombre d'attributs à renseigner : bien évaluer les besoins, prendre en compte ce qui peut être calculé (pour alléger l'encodage premier)
- Plusieurs couches d'encodage combinables (enrichissements progressifs)
- La compréhension fine du balisage d'un corpus / du système d'annotation permet de concevoir des interfaces adaptées à la question de recherche

Suite

- Le balisage n'est pas une opération mécanique et neutre, c'est une activité intellectuelle analytique orientée par des objectifs d'expression scientifique et des besoins d'exploitation
- La qualité du résultat dépend de la conscience que l'on a du code utilisé pour décoder, du système du nouveau codage et de la compréhension que l'on a de la façon dont les manipulations automatiques vont s'opérer
- C'est quelque chose qui peut faciliter le travail, faire gagner du temps et permettre une plus grande acuité scientifique dès le départ, même si on ne fait pas d'édition.

Conseils de lecture

- <http://www.pum.umontreal.ca/catalogue/pratiques-de-ledition-numerique>
- <http://books.openedition.org/oep/1298>
- <http://dh.obdurodon.org/what-is-xml.xhtml>
- <https://dlsanthology.commons.mla.org/digital-scholarly-editing/>
- <http://jtei.revues.org/>
- http://www.wwp.northeastern.edu/outreach/seminars/UCLA/presentations/html/introduction_markup_lecture.xhtml
- <http://mutec.huma-num.fr/sites/www.mutec-shs.fr/files/Guide%20Editer%20des%20sources%20historiques%20%20gr%C3%A2ce%20a%20XML.pdf>

