

# *Structuration du document avec XML*

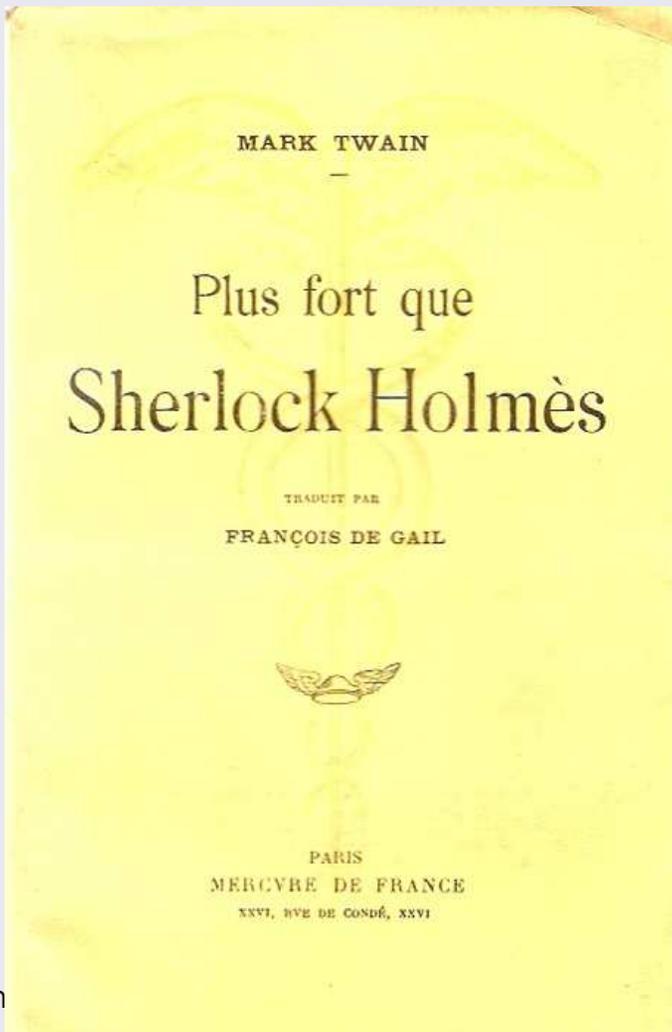
Bernard JACQUEMIN

Mulhouse – 14 décembre 2011

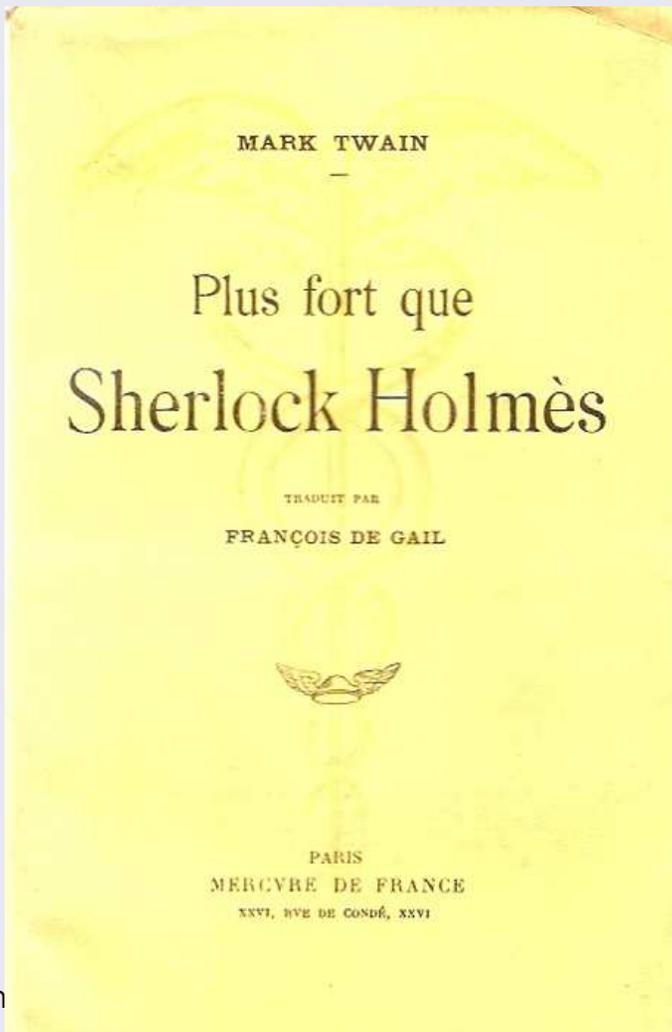
# Introduction

- XML permet d'identifier l'information (donner du **sens**)
- XML donne une **structure** à l'information
- XML peut intervenir directement **dans le document**

- XML permet d'identifier l'information (donner du **sens**)
- XML donne une **structure** à l'information
- XML peut intervenir directement **dans le document**

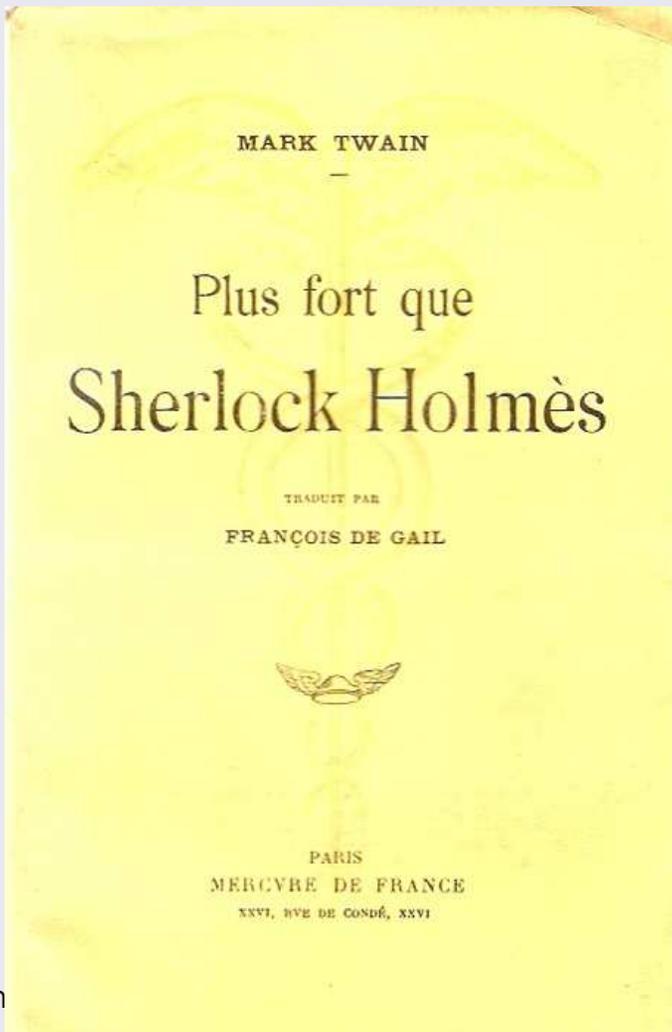


- XML permet d'identifier l'information (donner du **sens**)
- XML donne une **structure** à l'information
- XML peut intervenir directement **dans le document**



Identifiant pérenne de la notice :	<a href="http://www.sudoc.fr/049121111">http://www.sudoc.fr/049121111</a>
Titre :	<a href="#">Plus fort que Sherlock Holmès</a> / Mark Twain ; traduit par François de Gail. - 4e éd.
Auteur(s) :	<a href="#">Twain, Mark (1835-1910)</a> . Auteur <a href="#">Gail, François de</a> . Auteur
Date(s) :	1907
Langue(s) :	français
Pays :	France
Editeur(s) :	Paris : <a href="#">Société</a> du <a href="#">Mercure</a> de <a href="#">France</a> , 1907
Description :	223 p. ; 19 cm
Origine de la notice :	OCLC

- XML permet d'identifier l'information (donner du **sens**)
- XML donne une **structure** à l'information
- XML peut intervenir directement **dans le document**



Identifiant pérenne de la notice :	<a href="http://www.sudoc.fr/049121111">http://www.sudoc.fr/049121111</a>
Titre :	<a href="#">Plus fort que Sherlock Holmès</a> / Mark Twain ; traduit par François de Gail. - 4e éd.
Auteur(s) :	<a href="#">Twain, Mark (1835-1910)</a> , Auteur <a href="#">Gail, François de</a> , Auteur
Date(s) :	1907
Langue(s) :	français
Pays :	France
Editeur(s) :	Paris : <a href="#">Société du Mercure de France</a> , 1907
Description :	223 p. ; 19 cm
Origine de la notice :	OCLC

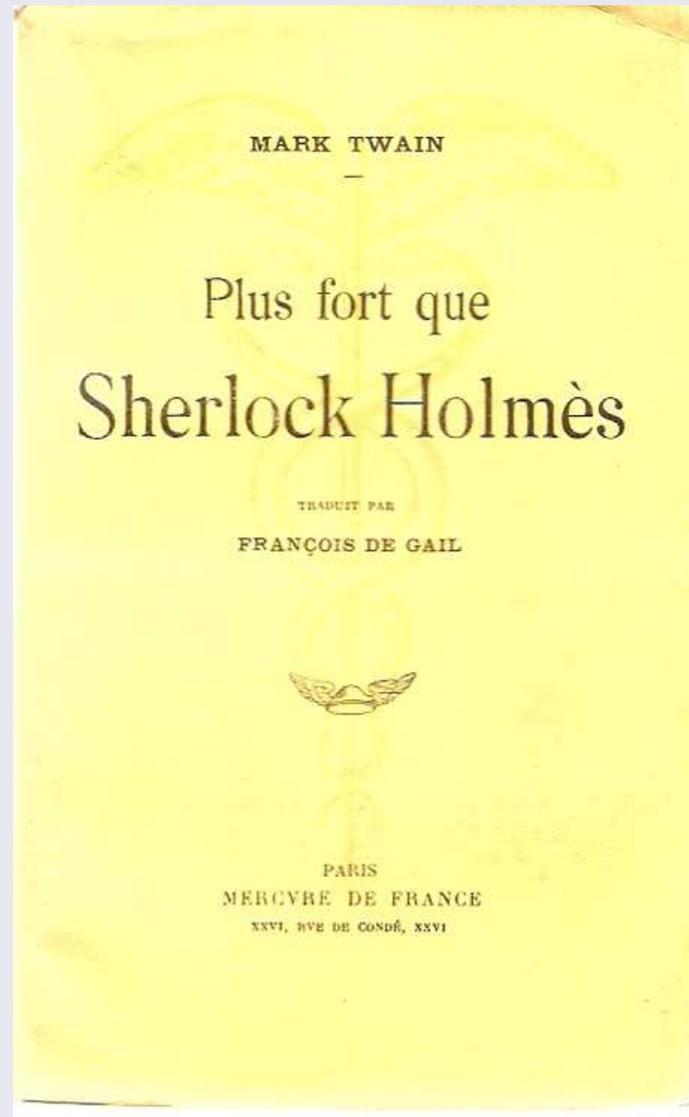
MARK TWAIN  
Plus fort que Sherlock Holmès  
TRADUIT PAR FRANÇOIS DE GAIL  
DEUXIÈME ÉDITION  
MCMVII  
PREMIÈRE PARTIE  
I  
La première scène se passe à la campagne dans la province de Virginie, en l'année 1880.  
Un élégant jeune homme de vingt-six ans, de fortune médiocre, vient d'épouser une jeune fille très riche. Mariage d'amour à première vue, précipitamment conclu, mais auquel le père de la jeune personne, un veuf, s'est opposé de toutes ses forces.

À quoi ressemble alors un document XML ?

Exemple du document précédent (téléchargé sur *Gutenberg Project*)

À quoi ressemble alors un document XML ?

Exemple du document précédent (téléchargé sur *Gutenberg Project*)



À quoi ressemble alors un document XML ?

Exemple du document précédent (téléchargé sur *Gutenberg Project*)

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>MARK TWAIN</auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>MARK TWAIN</auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>MARK TWAIN</auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>Mark Twain</auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traduction>TRADUIT PAR FRANÇOIS DE GAIL</traduction>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traducteur>FRANÇOIS DE GAIL</traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traducteur>François de Gail</traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que Sherlock Holmès</titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

... et d'autres organisations

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

... et d'autres organisations

... qui sont des **modélisations**

- XML = *eXtensible Markup Language*
  - un langage informatique... donc formel  
... mais il reste compréhensible et utilisable par un profane
  - un langage de balisage  
qui ajoute une information sur ce qu'il balise
  - un langage extensible
    - au niveau du vocabulaire
      - les balises ne sont **pas** prédéfinies
    - au niveau de la grammaire
      - la structure (DTD) est définie par l'utilisateur
      - grammaire = structure du document
      - grammaire = structure de l'information
  - une **modélisation** de l'information documentaire
  - pas de mise en forme directement attachée (CSS – XSL)

# XML et structure des contenus

- Utilisation de **balises** pour caractériser une information
- Balise = chaîne de caractères continue encadrée par < et >  
`<etablissement>`
- À toute balise ouvrante correspond une balise fermante  
`<statut>`toute chaîne de caractères, espace, tabulation, et même retours à la ligne sont autorisés`</statut>`
- On peut créer des balises sans contenu (*mile stone*)  
`< salaire />`
- Chevauchement de balises interdit  
`<page>texte1 <para>texte2</page>texte3</para>`  
`<page>texte1 <para>texte2</para></page>...`
- Une balise « principale » unique contient **tout** le document XML
- Prologue : `<?xml version="1.0" encoding="UTF-8" ?>`
- Notion : document bien formé (*well-formed*, « justesse de forme »)

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

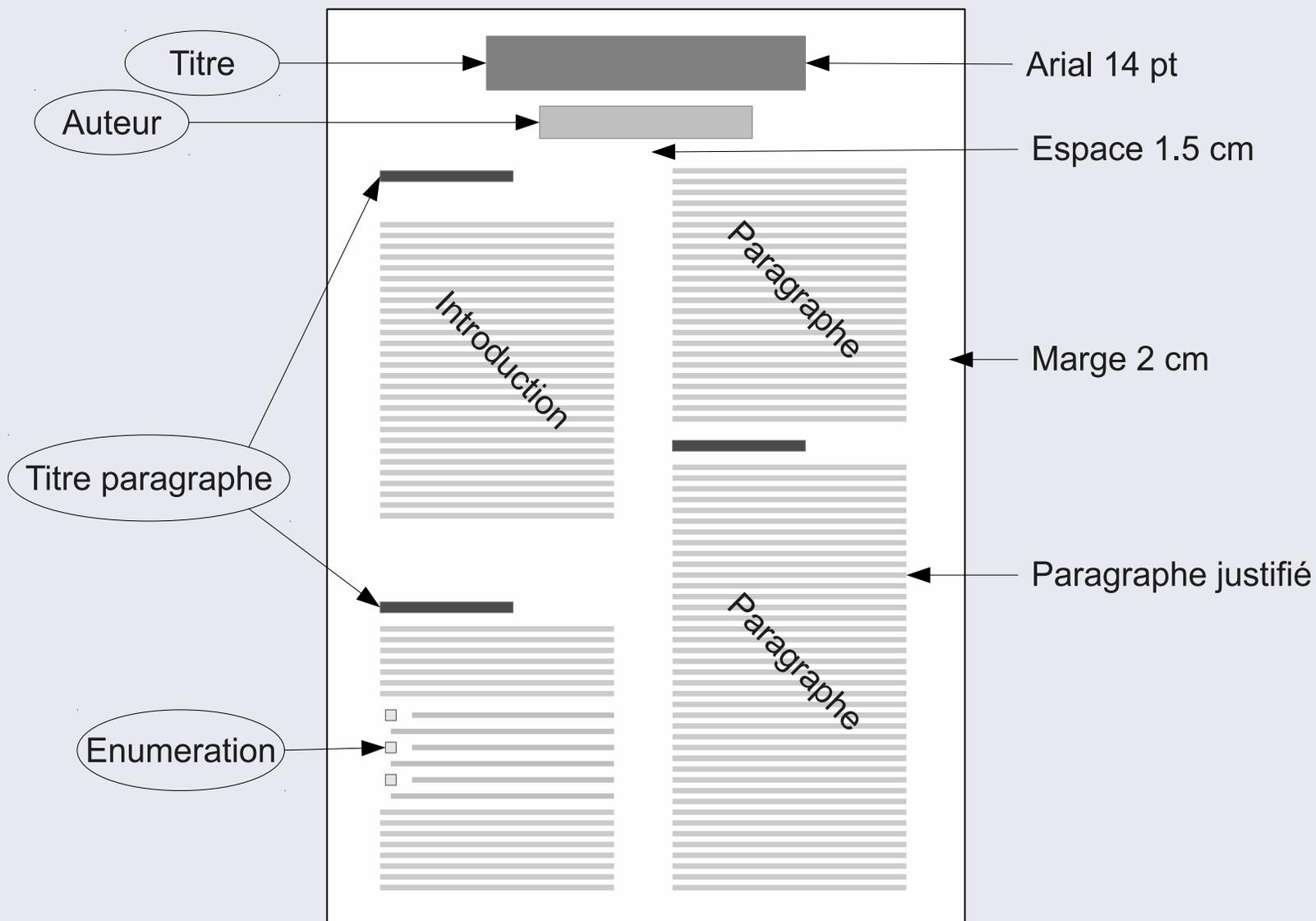
```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <prenom>Mark</prenom>
    <nom>Twain</nom>
  </auteur>
  <titre>Plus fort que <prenom>Sherlock</prenom> <nom>Holmès</nom></titre>
  <traducteur>
    <prenom>François</prenom>
    <nom>de Gail</nom>
  </traducteur>
  <edition>DEUXIÈME ÉDITION</edition>
  <annee>MCMVII</annee>
  <partie>PREMIÈRE PARTIE
    <chapitre>I
      <paragraphe>La première scène se passe à la campagne<finPage/> dans
        la province de Virginie, en l'année 1880.</paragraphe>
      <paragraphe>...</paragraphe>
      ...
    </chapitre>
  </partie>
</livre>
```

- Dans le XML
  - le balisage est **orienté contenu**
  - l'utilisateur peut créer ses propres balises

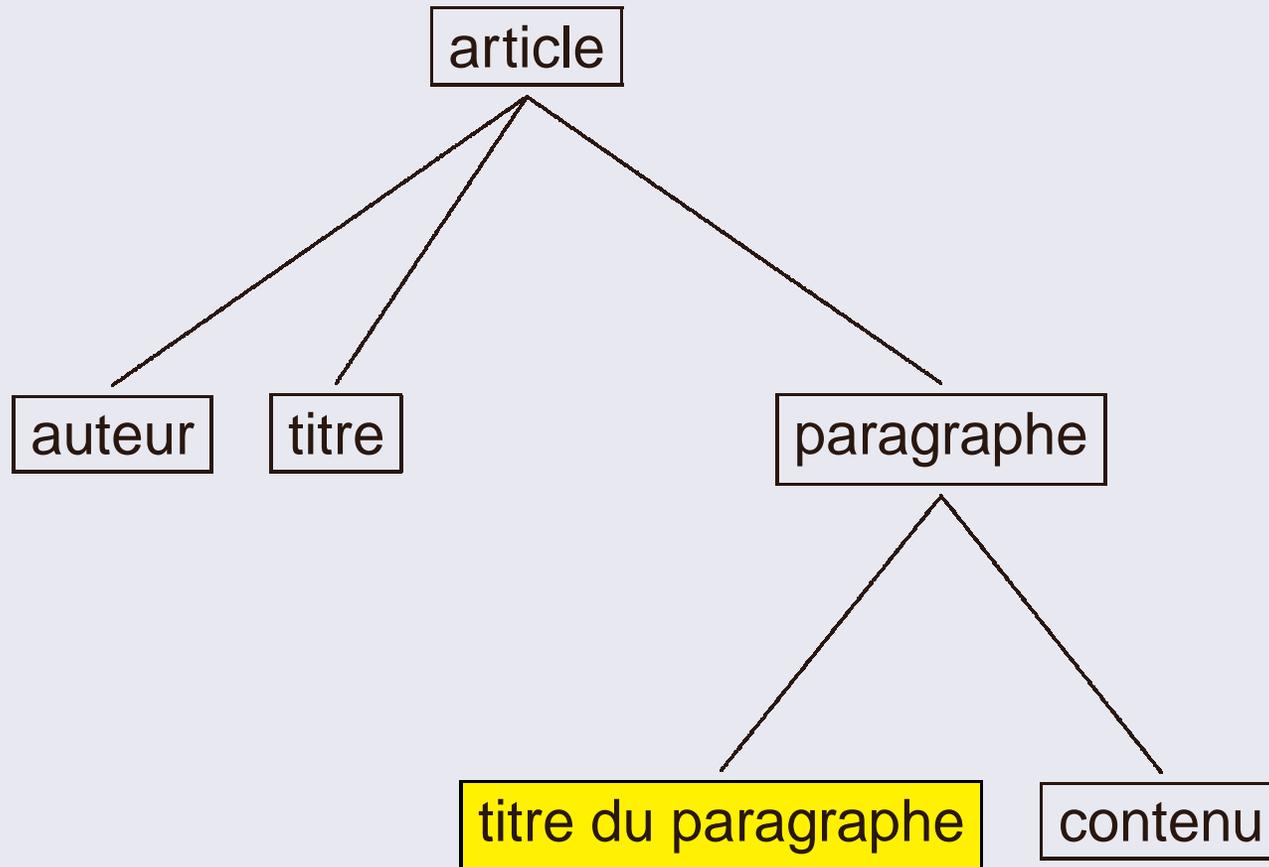
- Dans le XML
    - le balisage est **orienté contenu**
    - l'utilisateur peut créer ses propres balises
- ⇒ une même notion peut être décrite par plusieurs balises
- ⇒ réutilisation des balises d'autres utilisateurs

- Dans le XML
  - le balisage est **orienté contenu**
  - l'utilisateur peut créer ses propres balises
  - ⇒ une même notion peut être décrite par plusieurs balises
  - ⇒ réutilisation des balises d'autres utilisateurs ?
- Besoin d'une rationalisation dans la description des contenus
  - ⇒ formalisation/normalisation
  - ⇒ besoin d'un **modèle**
- Accès à un traitement logique de l'information via une modélisation qu'exploite XML

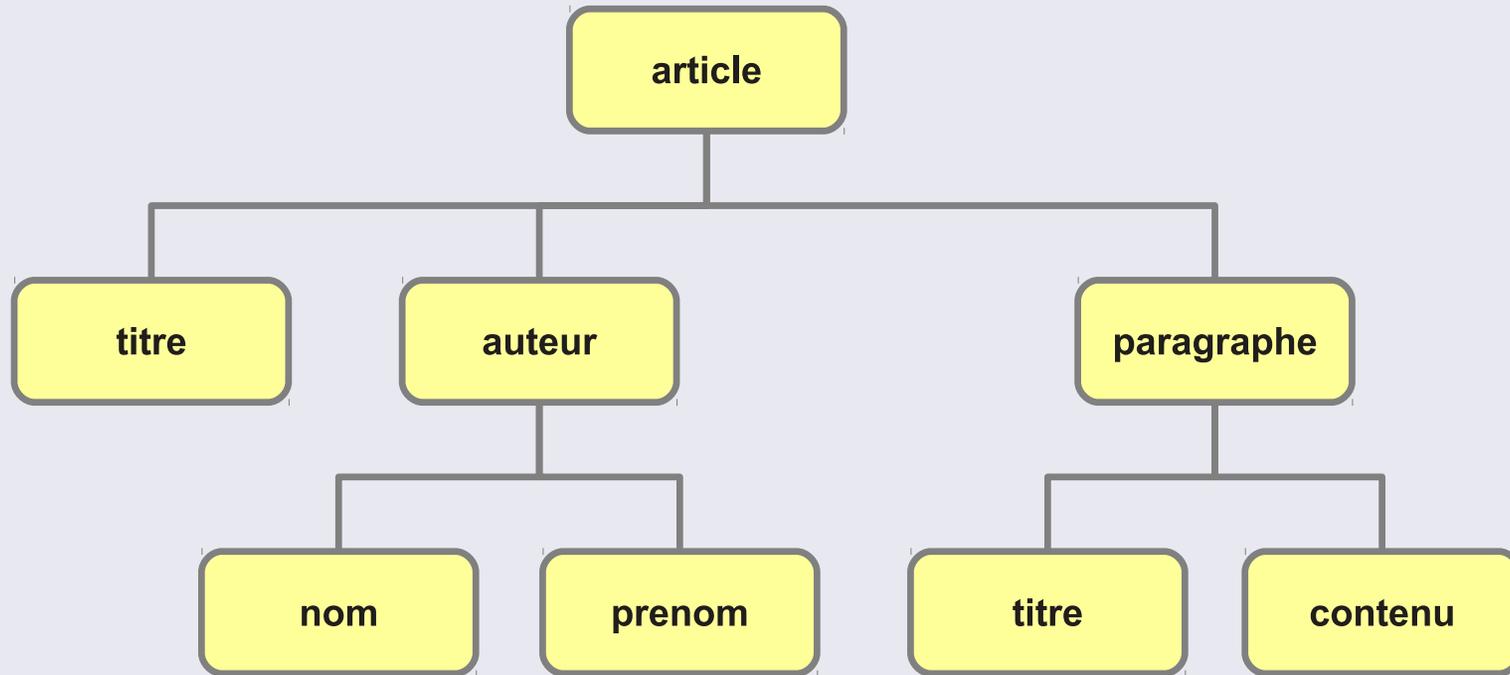
## Exemple d'un article rédigé et mis en forme sous Word



Exemple de **structuration** logique du document précédent

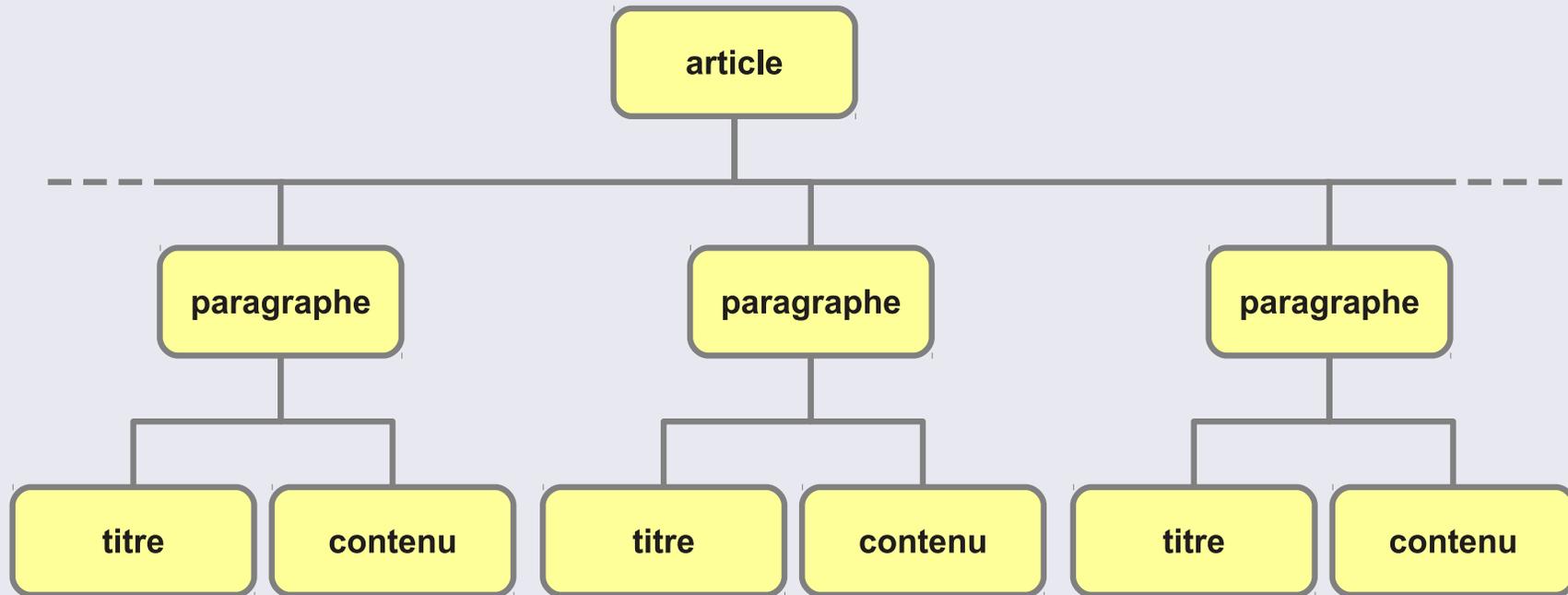


Version modélisée de la structure précédente

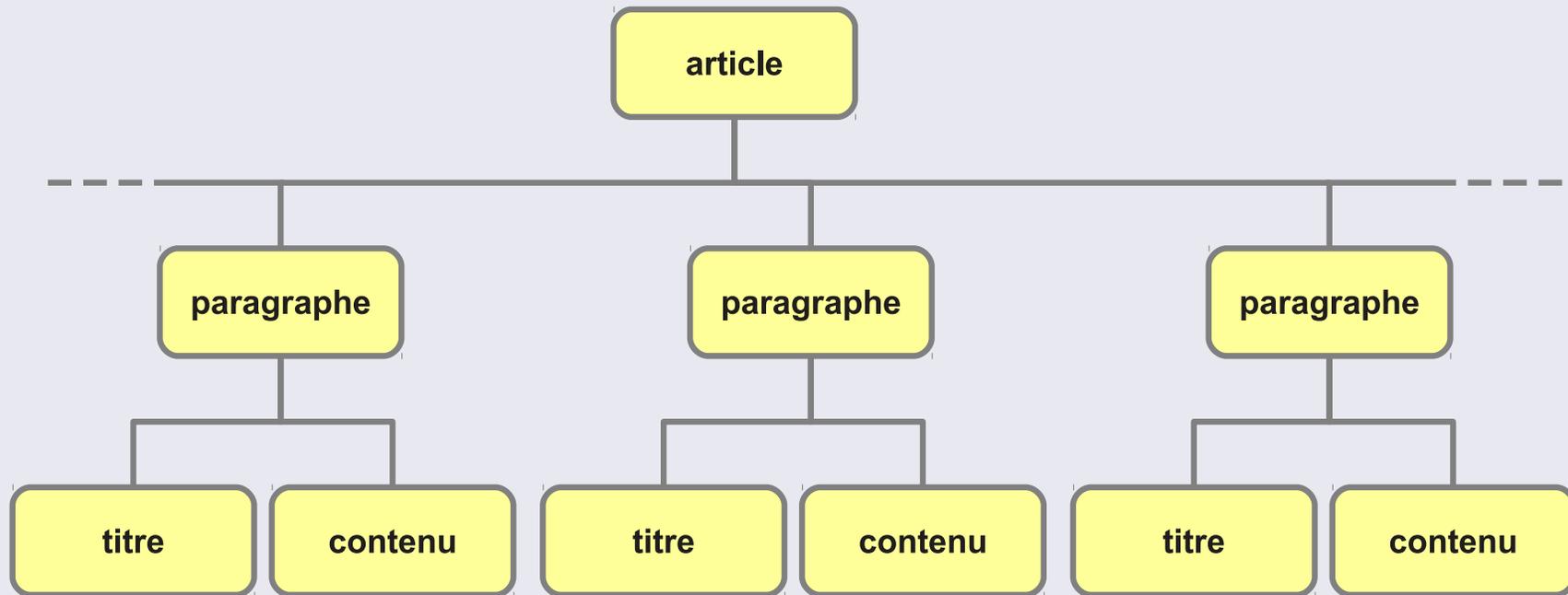


C'est un modèle déjà affiné : voir la déclinaison de l'auteur

À partir de la même structure de l'article, on pourrait proposer le formalisme (partiel) suivant :

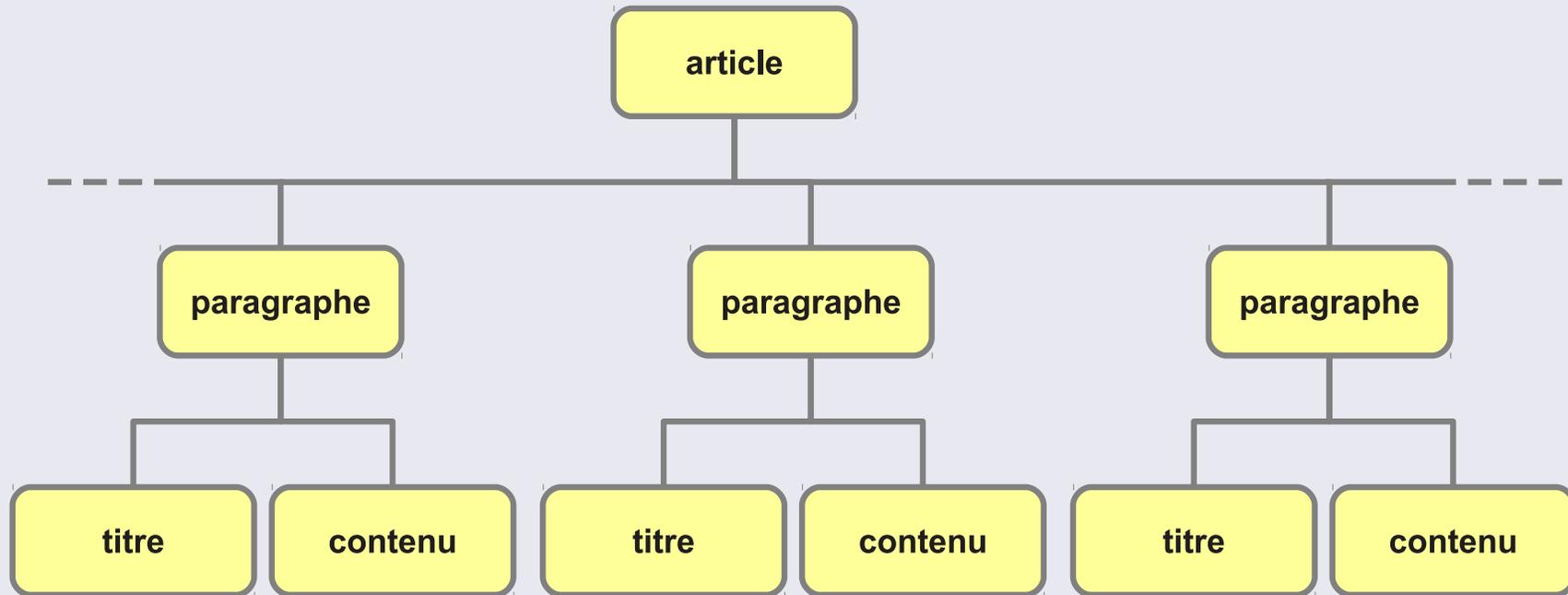


À partir de la même structure de l'article, on pourrait proposer le formalisme (partiel) suivant :



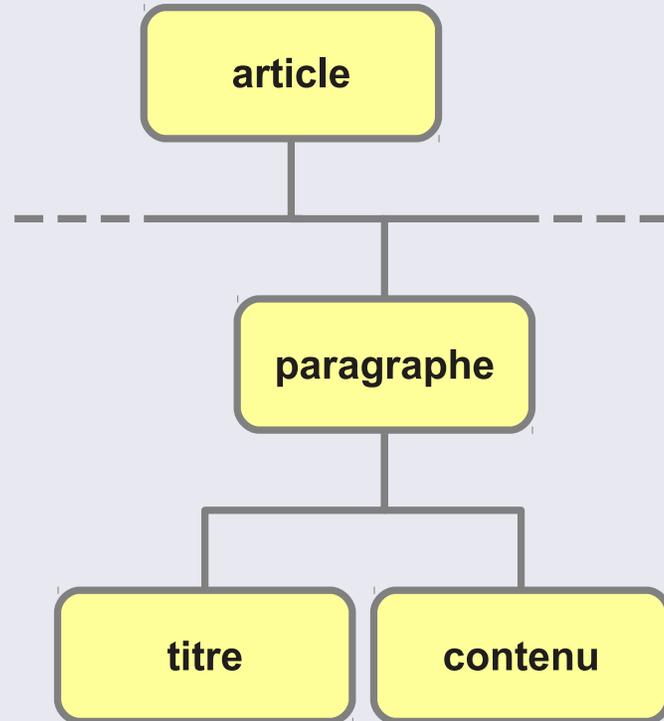
**Ce n'est pas un modèle !**

À partir de la même structure de l'article, on pourrait proposer le formalisme (partiel) suivant :



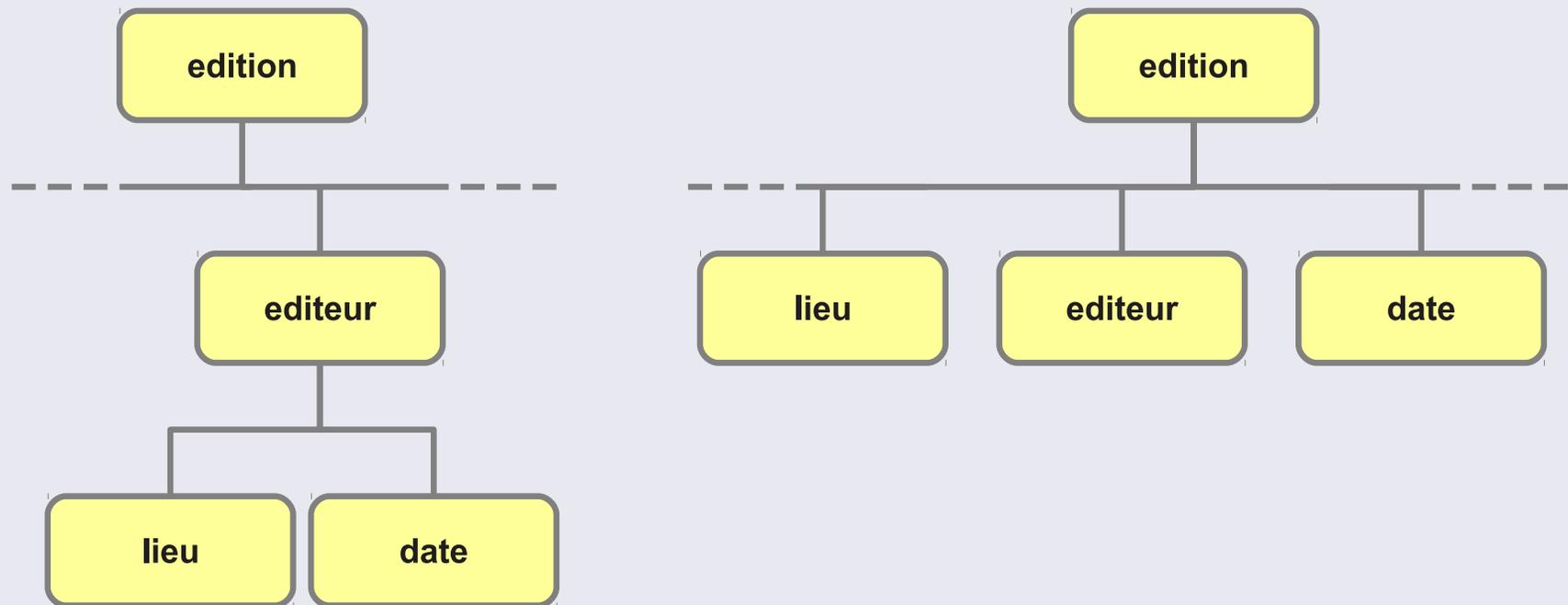
**Ce n'est pas un modèle !**

Dans un modèle (logique), une (sous-)branche n'est présente (représentée) qu'**une seule fois**



- cette **partie** de modèle est suffisante pour représenter la **structure** précédente dans le modèle de l'article
- la **représentation** du modèle n'indique pas s'il y a plusieurs paragraphes
- cette information est fournie par l'**explicitation informatique du modèle** : la **DTD** (*Document Type Definition*, à comprendre comme *Data Type Definition*)

- Un modèle est subjectif, orienté vers un usage, un objectif
- Plusieurs modèles peuvent être concurrents



- différences liées à la hiérarchie des branches
- pour un(e) documentaliste
  - quel choix ?
  - quels critères/besoins ?

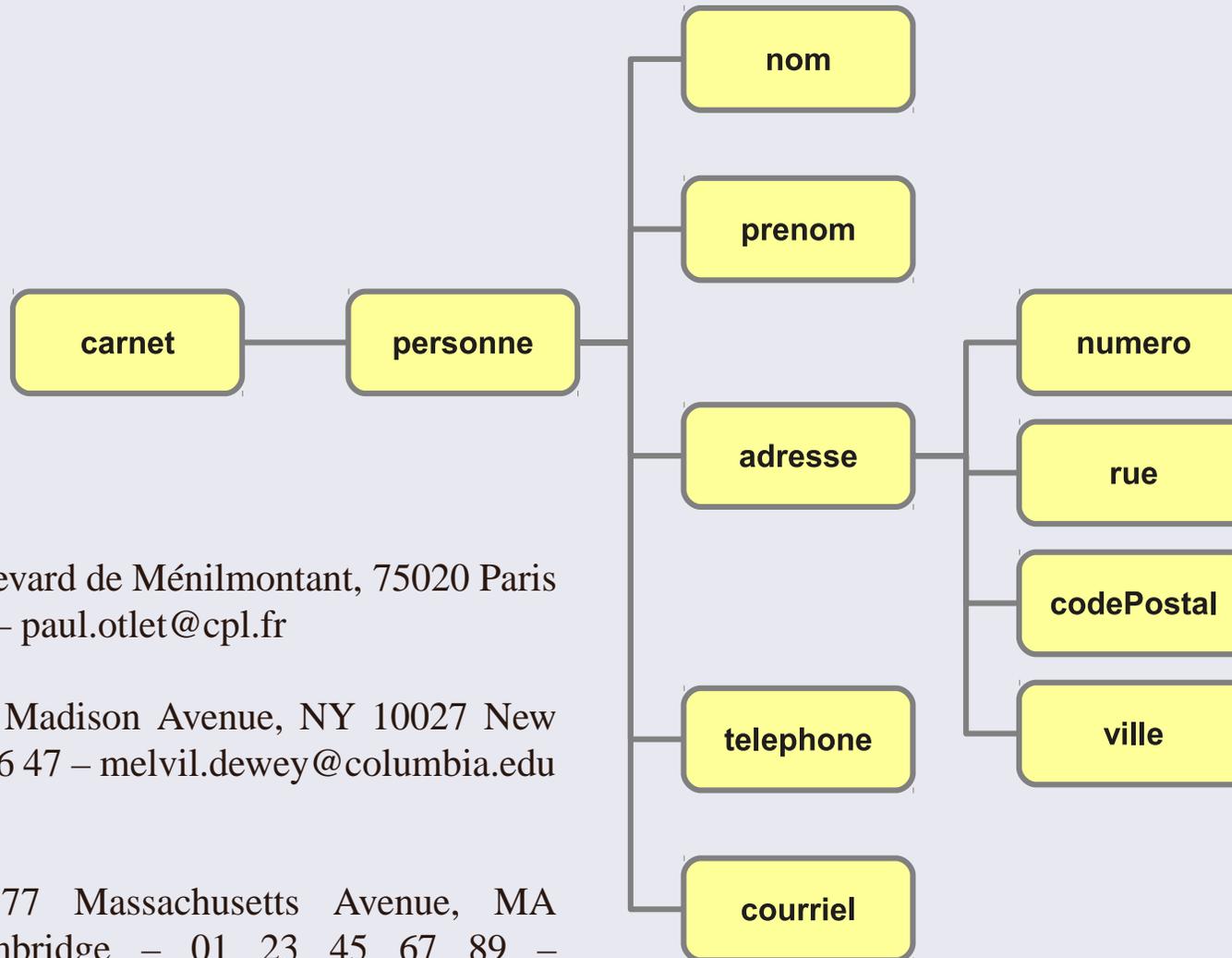
À partir des trois adresses présentées ci-dessous

- réaliser une structure arborescente de l'information proposée
  - créer un modèle sur base de la structure créée (éléments)
  - dans [XML Copy Editor](#), composer le document XML correspondant carnet1.xml (Dossier Carnet, sur le bureau)
- 
- Paul Otlet, 8 boulevard de Ménilmontant, 75020 Paris – 03 89 78 67 56 – paul.otlet@cpl.fr
  - Melvil Dewey, 1 Madison Avenue, NY 10027 New York – 06 14 25 36 47 – melvil.dewey@columbia.edu
  - Vanevar Bush, 77 Massachusetts Avenue, MA 02139-4307 Cambridge – 01 23 45 67 89 – vanevar.bush@rip.com

- On construit une **structure hiérarchique** (arborescence, *généalogie...*)
- L'arbre construit dispose d'**une seule racine**, qui ne sera **jamais dupliquée** (même dans le document XML)
- Les *autres* éléments sont dupliquables *dans le document XML* mais *pas dans le modèle*
- Deux types d'éléments (= étiquettes, types d'information)
  - étiquettes de structure (peu de contenu en dehors d'autres étiquettes)
  - étiquettes de contenu, qui reçoivent l'information (pas d'étiquettes « filles »)

- Pas de modèle idéal, absolu, objectif : **un modèle répond à un besoin**
- Principes de généalogie
  - un « enfant » ne peut avoir deux pères
  - différenciation de position dans l'arbre (emplacement des éléments)
    - rapport « père–fils » : hiérarchie
    - rapport entre « frères » : même niveau

## Modèle de carnet d'adresses (élaboré)



- Paul Otlet, 8 boulevard de Ménilmontant, 75020 Paris – 03 89 78 67 56 – paul.otlet@cpl.fr
- Melvil Dewey, 1 Madison Avenue, NY 10027 New York – 06 14 25 36 47 – melvil.dewey@columbia.edu
- Vanevar Bush, 77 Massachusetts Avenue, MA 02139-4307 Cambridge – 01 23 45 67 89 – vanevar.bush@rip.com

# La grammaire du XML : la DTD

- Type de document ?
  - **type de document : livre**
  - **type de document : article**
  - **type de document : carnet d'adresses**
- On s'intéresse à deux dimensions du document
  - les balises décrivant le contenu
  - l'agencement des balises (la structure du document)

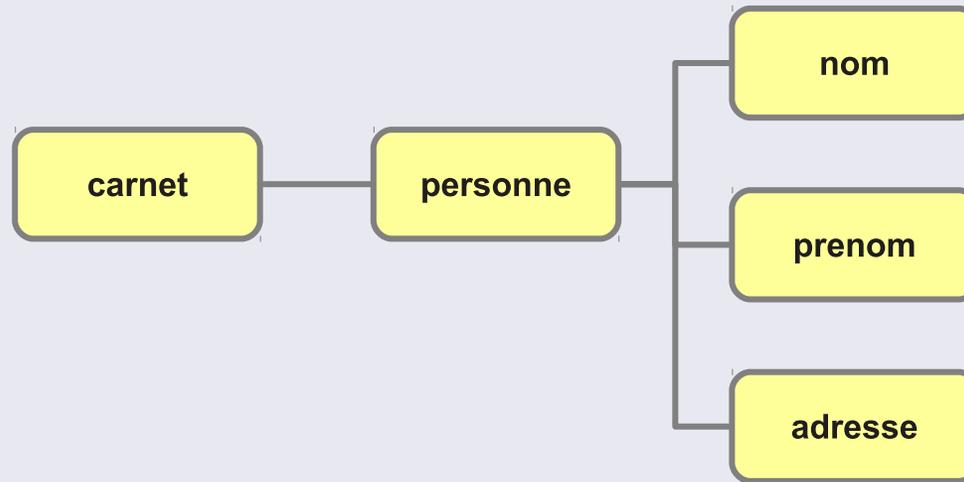
La Définition de Type de Document est la grammaire relative à un type de document XML

Grammaire = ensemble des règles qui régissent le fonctionnement d'une langue

- La Définition de Type de Document
  - décrit/représente le modèle hiérarchique dans un langage informatique ( $\neq$  XML)
  - prescrit la structure des documents XML liés (parfois réalisés directement dans le logiciel)
- Une DTD définit
  - la liste des balises (les différents éléments du modèle)
  - la possibilités (ou non) de dupliquer des balises dans le document
  - le type de contenu des balises (est-ce du texte ou un paramètre c.-à-d. une variable)
  - l'ordre de présentation des données (la hiérarchie)

- Un document XML doit en principe être
  - **bien formé** (notion de *well-formedness*) : respect de la syntaxe XML
  - **valide** (notion de *validity*) : respect de la DTD auquel il est attaché

DTD simplifiée du carnet d'adresses : carnet.dtd



`<!ELEMENT carnet (personne+) >`

`<!ELEMENT personne (nom, prenom, adresse) >`

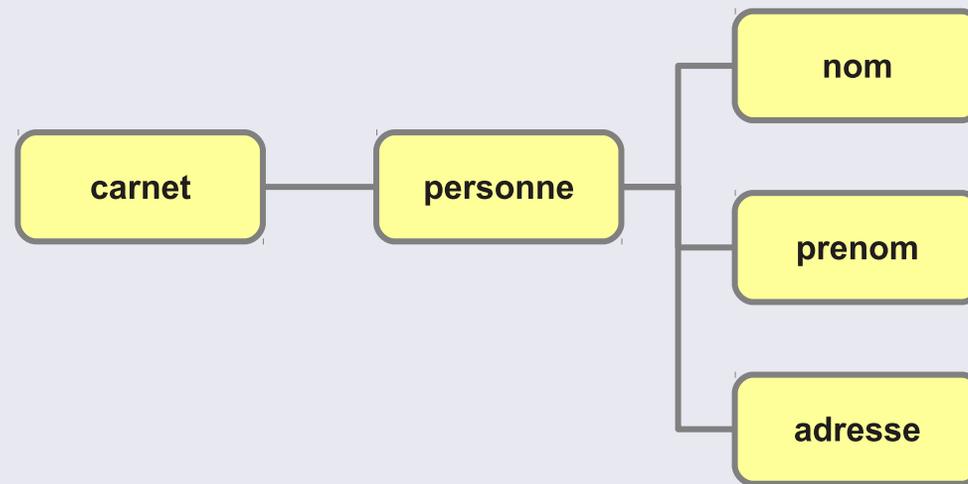
`<!ELEMENT nom (#PCDATA) >`

`<!ELEMENT prenom (#PCDATA) >`

`<!ELEMENT adresse (#PCDATA) >`

- Tout élément (étiquette) du modèle doit être déclaré  
`<!ELEMENT nomEtiquette ( contenu ) >`
- La composition de l'élément (contenu permis) doit être explicitée  
`<!ELEMENT nomEtiquette ( contenu ) >`
- Les contenus sont de deux types
  - indication d'élément(s)-fils : balise(s) dans l'elt courant  
`<!ELEMENT elementPere ( elementFils ) >`
  - indication d'information (contenus textuels)  
`<!ELEMENT nomEtiquette ( #PCDATA ) >`  
PCDATA = Parsed Character Data
- Les contenus sont quantifiables  
`<!ELEMENT elementPere ( elementFils+ ) >`  
+ signifie 1 ou plusieurs occurrences de elementFils

- Pratique sur le logiciel libre et gratuit [XML Copy Editor](#)
- Préparation : créer un répertoire « Carnet » *sur votre bureau*
- Lancer XML Copy Editor
- Créez une DTD « type » : Fichier > Nouveau... DTD (\*.dtd)
- À partir des deux lignes types proposées, écrire la DTD



- Enregistrez-la sous le nom `carnet.dtd` (menu Fichier > Enregistrer sous...) dans Carnet

- Créer un doc. XML : Fichier > Nouveau... Document XML (\*.xml)
- Un préambule (obligatoire) apparaît :  

```
<?xml version="1.0" encoding="UTF-8"?>
```
- Le contenu du document vient après le préambule
- Insérer l'étiquette de la racine selon la DTD `carnet.dtd` (`<carnet>`)
- Enregistrer le document XML sous le nom `carnet2.xml` (Fichier > Enregistrer sous...) dans `Carnet`
- Associer le document `carnet2.xml` à la DTD `carnet.dtd` : XML > Associer > DTD système... sélectionner `carnet.dtd`
- Rechargez les documents associés (Fichier > Reload)
- Complétez par les données de `carnet1.xml` (noms, prénoms, adresses) structurées selon la DTD (propositions conformes : insérer élément, insérer frère)

- Déclaration de tous les éléments (balises, étiquettes)

```
<!ELEMENT carnet (personne+)>
```

- Déclaration du contenu de chaque élément

- typage : autre(s) élément(s)

```
<!ELEMENT carnet (personne+)>
```

- typage : du texte (feuille, élément terminal)

```
<!ELEMENT nom (#PCDATA)>
```

- typage : texte ou n'importe quel élément

```
<!ELEMENT nom ANY>
```

- typage : élément vide

```
<!ELEMENT finPage EMPTY>
```

- Contenus d'éléments

- liste d'éléments obligatoires : opérateur « virgule »

```
<!ELEMENT personne (nom,prenom)>
```

- liste d'éléments alternatifs : opérateur « barre verticale »

```
<!ELEMENT contact (fixe|portable|email)>
```

- Quantification des contenus

- un ou plusieurs éléments : quantificateur « plus »

```
<!ELEMENT carnet (personne+)>
```

- élément optionnel (0 ou 1) : opérateur « point d'interrogation »

```
<!ELEMENT editeur (nom?,ville?)>
```

- élément optionnel répétable : opérateur « étoile »

```
<!ELEMENT reedition (annee*)>
```

- Téléchargez sur <http://jacqueminb.free.fr> l'exercice `recettes1.xml` (mal formé)
- Ouvrez-le dans [\*\*XML Copy Editor\*\*](#)
- Corrigez-en la syntaxe pour qu'il soit bien formé
- Construisez-lui une DTD `recettes.dtd`
- Associez `recettes1.xml` et `recettes.dtd` et rechargez
- Vérifiez la validité de `recettes1.xml`
- Contrôlez la pertinence de votre DTD :
  - téléchargez `recettes2.xml` sur le même site (bien formé)
  - associez `recettes2.xml` et `recettes.dtd` et rechargez
  - vérifiez la validité de `recettes2.xml` (modifiez la DTD...)

# Présentation XML via CSS

- Fichier XML : structuration des données sans préoccupation de mise en forme
- Fichier DTD : modèle conditionnant la structure de l'information, grammaire
- Fichier CSS : feuille de style capable de déterminer l'affichage des balises XML dans un navigateur
  - CSS : *Cascading StyleSheet*
  - langage informatique associant balise et mise en forme
  - association possible à plusieurs fichiers
- XSL : langage de feuilles de style *complexe* associé à XML  
(XSLT, XPath, XSL-FO)

- Les règles de mise en forme CSS sont associées à des balises XML
- Ces règles se trouvent dans un fichier spécifique (extension `.css`)
- Un fichier `.css` n'est pas du XML
- Association fichier XML – sa feuille de style *dans le fichier XML*  
`<?xml-stylesheet type="text/css" href="fichierDeRègles.css"?>`
- Pas de feuille CSS interne ( $\neq$  HTML)
- Syntaxe-type des règles CSS

```
balise {  
    prop1:val1;  
    prop2:val2;  
    ...  
}
```

```
sélecteur {  
    propriété :valeur; [déclaration]  
    propriété :valeur;  
    ...  
}
```

- Règle CSS à *sélecteur simple* : appliquée à l'élément

```
nom {  
    font-weight:bold;  
    font-size:10pt;  
}
```

- Règle CSS pour *liste de sélecteurs* : appliquée à tous les éléments

```
nom, prenom {  
    font-weight:bold;  
    font-size:10pt;  
}          nom {  
          font-variant:small-caps;  
          color:red;  
}
```

- Règle CSS pour *sélecteur contextuel* : élément d'un autre élément

```
paragraphe titre {  
    font-style:italic;  
    font-size:12pt;  
}          titre {  
          font-weight:bold;  
          font-size:14;  
}
```

- Propriété CSS pour l'affichage d'un élément : **display**
  - valeur **block** : affichage de l'élément sur une ligne propre  

```
livre { display:block; }
```
  - ▷ valeur **inline** : affichage de l'élément sur la même ligne  

```
livre { display:inline; }
```
  - valeur **none** : l'élément n'est pas affiché  

```
livre { display:none; }
```
  - valeur **list-item** : chaque élément est un item de liste  

```
livre { display:list-item; }
```

Pas d'héritage de déclaration pour la propriété *display*

- Notion : le commentaire en CSS « /\* \*/ » (pour test)

```
livre {  
    /* display:none; */  
    display:block;  
}
```

- Dans le répertoire `Carnet` sur votre bureau

- créez un fichier `carnet2.css` (vide)
- ajoutez-y les règles CSS suivantes :

```
personne {
    display: list-item;
}
adresse {
    display: none;
}
```

- Copiez le fichier XML `carnet2.xml` dans un fichier `carnet3.xml`

- visualisez-le avec un navigateur (IE, Firefox...)
- ajoutez sous le prologue la ligne suivante

```
<?xml-stylesheet type="text/css" href="carnet2.css"?>
```

- actualisez la visualisation dans le navigateur
- modifiez les propriétés d'affichage dans la règle CSS et visualisez-en les effets

- Modifiez la feuille CSS `carnet2.css`
  - pour que `carnet3.xml` se présente comme suit  
Nom prénom  
Numéro et rue  
Code Postal Ville  
Tel  
(espacement 0,7 cm)
- Problèmes de conception ! Réfléchir à la structure *en modélisant*
- Tentez de réaliser cet affichage avec le fichier `carnet1.xml`  
(pensez à associer `carnet2.css` et `carnet1.xml`)

- Propriétés CSS pour la mise en forme de type « page »
  - **margin** : réglage des « marges » (bords du navigateur)
    - `contact { margin-left:15mm; }` : gauche (en millimètres)
    - `contact { margin-top:40pt; }` : haut (en points)
    - `contact { margin-right:5em; }` : droite (en largeur de M)
    - `contact { margin-bottom:5ex; }` : bas (en hauteur de x)
    - `contact { margin:8%; }` : toutes marges (en % du total)
  - **border** : bordures, cadres autour de l'élément sélecteur
    - `contact { border-style:solid; }` : bordure continue  
autres valeurs courantes : *dotted* (pointillés), *double*, *dashed* (tirets)
    - `contact { border-width:0.5ex; }` : épaisseur du trait
    - `contact { border-color:red; }` : couleur du trait  
16 valeurs de base prédéfinies + valeurs RGB
  - **background-color** : couleur de fond de l'élément sélecteur
    - `contact { background-color:teal; }` : couleur du fond  
16 valeurs de base prédéfinies + valeurs RGB

- Couleurs prédéfinies pour le CSS

mots-clefs		Correspondances RGB	
	aqua	rgb(0,255,255)	rgb(0%,100%,100%)
	black	rgb(0,0,0)	rgb(0%,0%,0%)
	blue	rgb(0,0,255)	rgb(0%,0%,100%)
	fuchsia	rgb(255,0,255)	rgb(100%,0%,100%)
	gray	rgb(128,128,128)	rgb(50%,50%,50%)
	green	rgb(0,128,0)	rgb(0%,50%,0%)
	lime	rgb(0,255,0)	rgb(0%,100%,0%)
	maroon	rgb(128,0,0)	rgb(50%,0%,0%)
	navy	rgb(0,0,128)	rgb(0%,0%,50%)
	olive	rgb(128,128,0)	rgb(50%,50%,0%)
	purple	rgb(128,0,128)	rgb(50%,0%,50%)
	red	rgb(255,0,0)	rgb(100%,0%,0%)
	silver	rgb(192,192,192)	rgb(75%,75%,75%)
	teal	rgb(0,128,128)	rgb(0%,50%,50%)
	white	rgb(255,255,255)	rgb(100%,100%,100%)
	yellow	rgb(255,255,0)	rgb(100%,100%,0%)

- Propriétés CSS pour la mise en forme de type « paragraphe »
  - **text-align** : alignement du texte sur les marges
    - `contact { text-align: justify; }` : aligné à gauche et à droite
    - ▷ `contact { text-align: left; }` : aligné à gauche
    - `contact { text-align: right; }` : aligné à droite
    - `contact { text-align: center; }` : centré
  - **text-indent** : retrait (indentation) de la *1<sup>re</sup> ligne* de l'élément  
`contact { text-indent: 5em; }` : indentation relative de 5 M
  - **line-height** : taille de l'interligne (qq soit la taille des caractères)  
`contact { line-height: 2em; }` : interligne  $2em = 200\% \approx 24pt$

- Propriétés CSS pour la mise en forme de type « police »
  - **font-family** : indique la police ou la famille de police à utiliser

contact { **font-family:Times,serif;** } : police Times, ou serif

mots-clefs génériques	noms de polices
<i>serif</i>	Times, "Times New Roman", Palatino, ...
<i>sans-serif</i>	Arial, Helvetica, Geneva, Verdana, ...
<i>cursive</i>	ZapfChancery, ...
<i>fantasy</i>	Western, Critter, ...
<i>monospace</i>	Courier, System, ...

- **font-size** : taille de police relative ou absolue (pt)

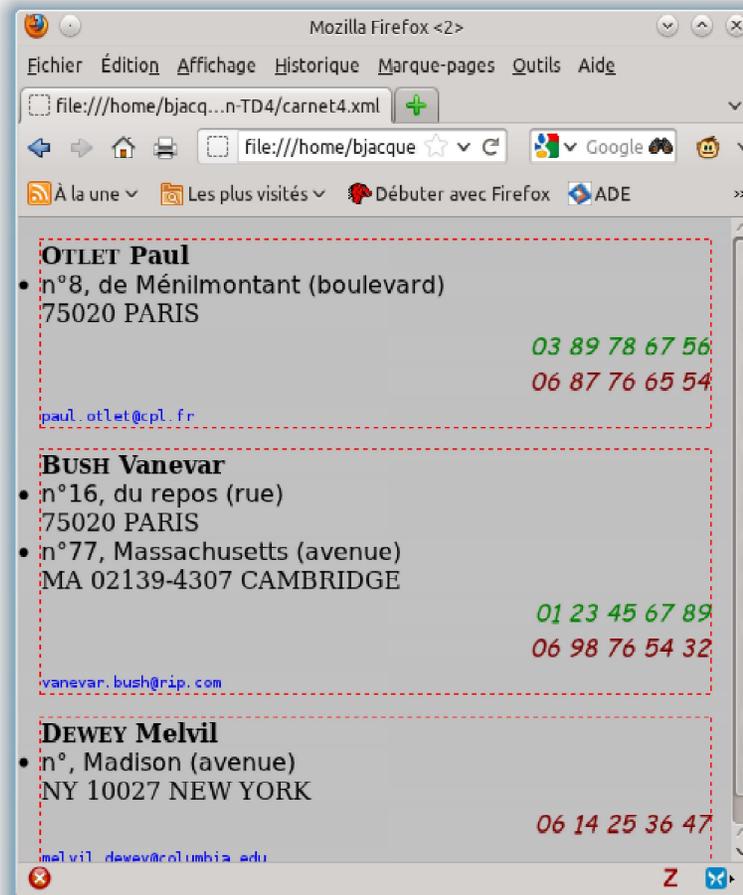
contact { **font-size:11pt;** } : taille de police = 11pt

Valeurs relatives à la taille de police du navigateur :	
<i>xx-small</i>	la taille la plus petite possible
<i>x-small</i>	1.5 fois la taille de <i>xx-small</i>
<i>small</i>	1.5 fois la taille de <i>x-small</i>
<b><i>medium</i></b>	1.5 fois la taille de <i>small</i>
<i>large</i>	1.5 fois la taille de <i>medium</i>
<i>x-large</i>	1.5 fois la taille de <i>large</i>
<i>xx-large</i>	1.5 fois la taille de <i>x-large</i>

Valeurs relatives à la taille de police de l'élément parent :	
<i>smaller</i>	33% plus petit que la police de l'élément parent
<i>larger</i>	50% plus grand que la police de l'élément parent

- Propriétés CSS pour la mise en forme de type « police »
  - **color** : couleur de police (16 couleurs + valeurs RGB)  
`contact { color:lime; }`
  - **font-style** : style de police (*oblique*, *italic*, *normal*)  
`contact { font-style:oblique; }` : caractères penchés
  - **font-weight** : graisse de la police (*bold*, *bolder*, *lighter*, *normal*)  
`contact { font-weight:bolder; }` : caractères plus gras
  - **text-decoration** : (sou)lignement (*underline*, *overline*, *line-through*)  
`contact { text-decoration:line-through; }` : caractères barrés
  - **font-variant** : police en petites capitales (*small-caps*, *normal*)  
`contact { font-variant:small-caps; }` : petites capitales
  - **text-transform** : casse de police (*uppercase*, *capitalize*, *lowercase*)  
`contact { font-weight:capitalize; }` : 1<sup>re</sup> lettre des mots majuscule

- À partir du fichier `carnet1.xml` déjà utilisé :
  - Construisez une feuille de style `carnet4.css` qui lui est attachée (en remplacement)
  - Créez les règles qui permettent la mise en forme de la figure ci-dessous



# XML avancé : attributs

- Vous connaissez déjà l'**élément**

`< adresse >16 rue de la Fonderie< /adresse >`

- L'information de l'élément peut être précisée par un **attribut**

`< adresse type = "professionnelle" >`

- l'attribut est composé d'un nom et d'une valeur (guillemets)

- l'attribut n'apparaît que dans la balise ouvrante

`< adresse type = "pro" >16 rue de la Fonderie< /adresse >`

- un élément peut contenir un ou plusieurs attributs

`< adresse type = "perso" pays = "france" >`

- un attribut ne peut apparaître qu'une fois dans le même élément

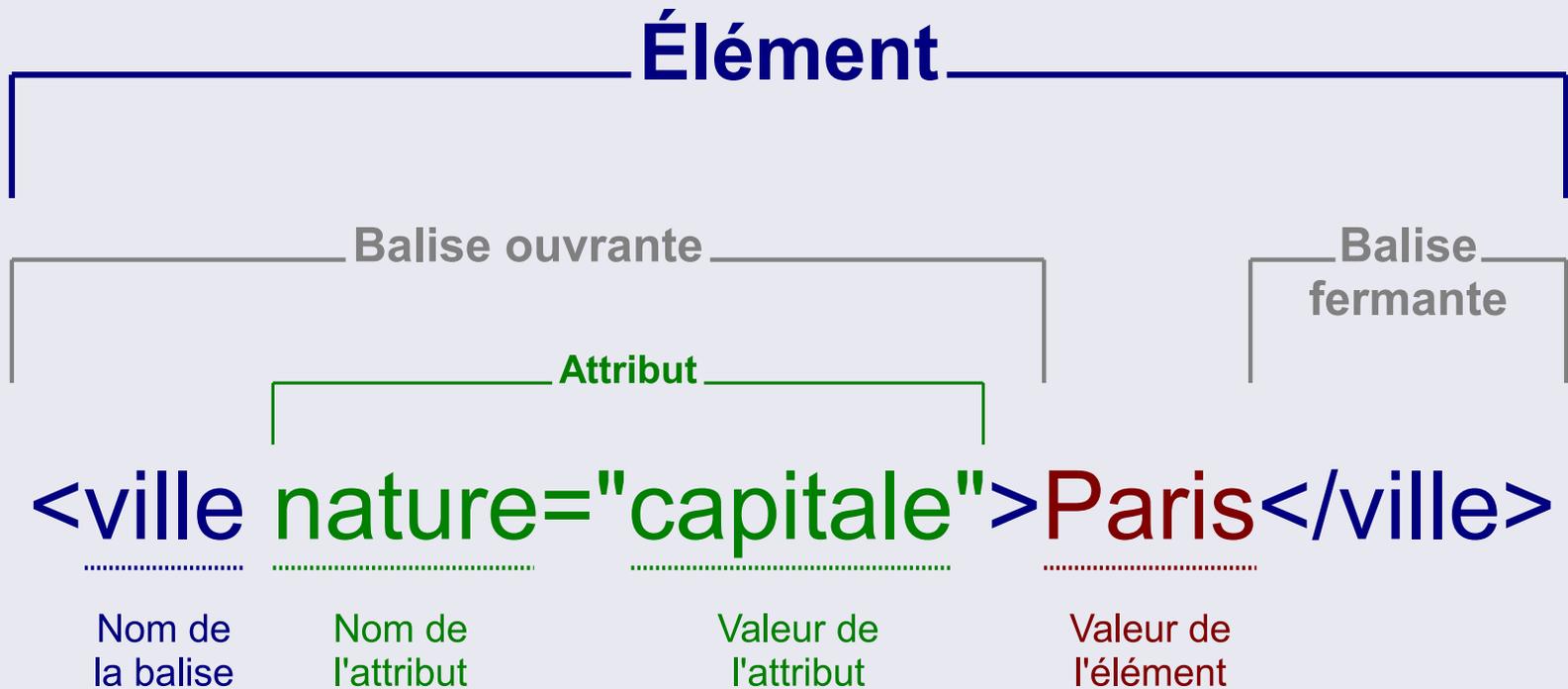
- Intérêt de l'attribut

- pour spécifier différentes informations de même nature

- pour éviter de multiplier inutilement les éléments

voir `< quantite unite = "kg" >` vs `< quantite unite = "CS" >`

- Vocabulaire : éléments et attributs



- Vocabulaire : éléments et attributs
- Un exemple connu : carnet d'adresses

```
<carnet>
  <contact>
    <identite><prenom>Paul</prenom><nom>Otlet</nom></identite>
    <telephone>03 89 78 67 56</telephone>
    <portable>06 87 76 65 54</portable>
    ...
  </contact>
  <contact>
    <identite><prenom>Vanevar</prenom><nom>Bush</nom></identite>
    <telephone>01 23 45 67 89</telephone>
    <portable>06 98 76 54 32</portable>
    ...
  </contact>
</carnet>
```

- Vocabulaire : éléments et attributs
- Un exemple connu : carnet d'adresses

```
<carnet>
  <contact>
    <identite><prenom>Paul</prenom><nom>Otlet</nom></identite>
    <tel type="fixe">03 89 78 67 56</tel>
    <portable>06 87 76 65 54</portable>
    ...
  </contact>
  <contact>
    <identite><prenom>Vanevar</prenom><nom>Bush</nom></identite>
    <telephone>01 23 45 67 89</telephone>
    <portable>06 98 76 54 32</portable>
    ...
  </contact>
</carnet>
```

- Vocabulaire : éléments et attributs
- Un exemple connu : carnet d'adresses

```
<carnet>
  <contact>
    <identite><prenom>Paul</prenom><nom>Otlet</nom></identite>
    <tel type="fixe">03 89 78 67 56</tel>
    <tel type="mobile">06 87 76 65 54</tel>
    ...
  </contact>
  <contact>
    <identite><prenom>Vanevar</prenom><nom>Bush</nom></identite>
    <telephone>01 23 45 67 89</telephone>
    <portable>06 98 76 54 32</portable>
    ...
  </contact>
</carnet>
```

- Vocabulaire : éléments et attributs
- Un exemple connu : carnet d'adresses

```
<carnet>
  <contact>
    <identite><prenom>Paul</prenom><nom>Otlet</nom></identite>
    <tel type="fixe">03 89 78 67 56</tel>
    <tel type="mobile">06 87 76 65 54</tel>
    ...
  </contact>
  <contact>
    <identite><prenom>Vanevar</prenom><nom>Bush</nom></identite>
    <tel type="fixe">01 23 45 67 89</tel>
    <portable>06 98 76 54 32</portable>
    ...
  </contact>
</carnet>
```

- Vocabulaire : éléments et attributs
- Un exemple connu : carnet d'adresses

```
<carnet>
  <contact>
    <identite><prenom>Paul</prenom><nom>Otlet</nom></identite>
    <tel type="fixe">03 89 78 67 56</tel>
    <tel type="mobile">06 87 76 65 54</tel>
    ...
  </contact>
  <contact>
    <identite><prenom>Vanevar</prenom><nom>Bush</nom></identite>
    <tel type="fixe">01 23 45 67 89</tel>
    <tel type="mobile">06 98 76 54 32</tel>
    ...
  </contact>
</carnet>
```

- La DTD structure l'information XML
  - grammaire des **éléments**
  - grammaire des **attributs** d'éléments
- Les attributs d'un élément doivent être déclarés *pour cet élément*  
`<!ATTLIST quantite unite CDATA "litre" >`
  - les valeurs d'attributs sont **typées**  
`<!ATTLIST elt attr type option >`
    - CDATA : la valeur d'attribut est une chaîne de caractères *indéterminée*
    - ID : la valeur d'attribut est une chaîne de caractère *unique* (identifiant)
    - (val1 | val2) : la valeur d'attribut est dans la liste (énuméré)

- Les attributs d'un élément doivent être déclarés *pour cet élément*  
`<!ATTLIST quantite unite CDATA "litre" >`
- l'utilisation d'un attribut est spécifiée par une **option**  
`<!ATTLIST elt attr type option >`
  - `#IMPLIED` : l'utilisation de l'attribut est optionnelle
  - `#REQUIRED` : l'utilisation de l'attribut est obligatoire
  - `"val1"` : l'attribut optionnel est de valeur `val1` par défaut
  - `#FIXED "val1"` : l'attribut obligatoire est toujours de valeur `val1` (compatibilité descendante de DTD)

- Exemple connu : DTD du carnet d'adresses
  - téléchargez sur <http://jacqueminb.free.fr> le fichier `carnetAttributs.xml` dans un dossier `Attributs`
  - créez un fichier de DTD `carnetAttributs.dtd` dans le même dossier
  - associez la DTD `carnetAttributs.dtd` au fichier XML `carnetAttributs.xml`
  - concevez la DTD décrivant les éléments et les attributs du carnet d'adresses
    - type d'adresse par défaut professionnelle
    - type de rue : type énuméré

- Règle CSS en fonction d'un attribut de l'élément sélecteur

```
livre[edEbook] {  
    font-family:serif;  
    color:blue;  
}
```

s'applique si `edEbook` est présent

```
livre[edEbook="Garon"] {  
    font-weight:bold;  
    color:silver;  
}
```

s'applique si `edEbook` est *Garon*

- Règle CSS permettant d'afficher l'attribut d'un sélecteur

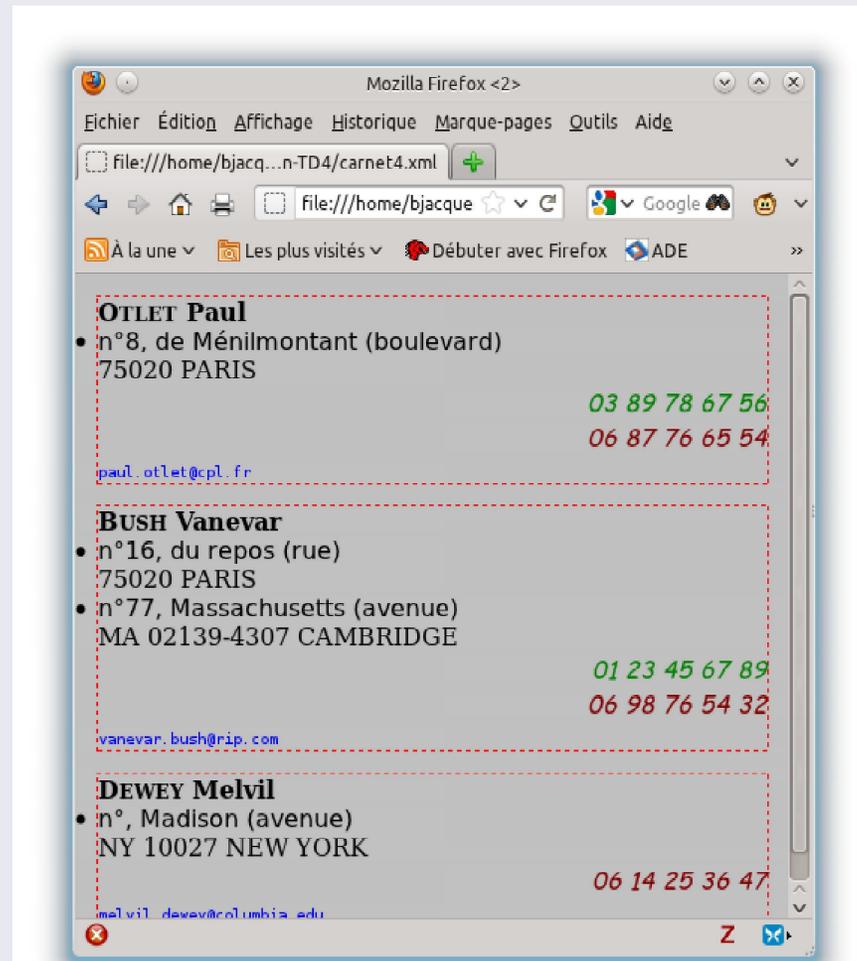
```
livre:before {  
    content:"l'éditeur "  
    attr(edEbook)  
    " publie " ;  
}
```

pour « l'éditeur **Garon** publie  
*titreLivres* »

```
livre:after {  
    content:", publié par "  
    attr(edEbook) ;  
}
```

pour « *titreLivres*, publié par **Garon** »

- Téléchargez le fichier `carnet4.xml` déjà utilisé :
- Construisez une feuille de style `carnet5.css` qui lui est attachée
- Créez les règles qui permettent la mise en forme de la figure ci-dessous



- À partir de la bibliographie ci-dessous (pas les ponctuations)
  - créez un modèle hiérarchique de bibliographie
  - dans un dossier `Biblio`, créez-en la DTD `biblio.dtd` (avec attributs **nécessaires**)
  - établissez le fichier `biblio.xml` bien formé et valide
  - construisez un fichier `biblio.css` permettant d'obtenir l'affichage correspondant

- **DE CERTEAU, Michel.** *L'invention du quotidien. 1. Arts de faire.* Paris : Gallimard, 1980.
- **VIÉGAS, Fernanda B., WATTENBERG, Martin, DAVE, Kushal.** Studying Cooperation and Conflict between Authors with history flow Visualizations. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienne, 2004, p. 575–582.
- **ZLATIC, Vinko, BOŽIČEVIĆ, Miran, ŠTEFANČIĆ, Hrvoje, DOMAZET, Mladen.** Wikipedias : Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 2006, 74, 1, p. 6–11.