

Objectif

- Le balisage sémantique
- Les entités nommées : qu'est-ce que c'est ?
- Comment les encoder ?
 - Le module `namesdates`
 - Découvrir les éléments de la TEI pour indexer les entités nommées

Balisage sémantique

- Première étape du balisage :
 - Structurel et formel
- Seconde étape du balisage :
 - Sémantique
- Troisième étape du balisage : l'analyse interprétative du contenu du texte
 - <seg> avec @ana
 - <interpGrp> et <interp>

Balisage sémantique

- Définition résolument informatique :
 - A la différence de l'analyse lexicale qui s'occupe d'un mot dans une phrase, l'analyse sémantique s'intéresse à la signification du mot (pas du contexte) : le concept
 - Au niveau lexical, la Tour Eiffel = 2 mots alors qu'au niveau sémantique = 1 entité nommé
 - Il autorise un encodage d'un premier niveau d'analyse comme les noms, les toponymes et la temporalité

La temporalité

- Les éléments :
 - <date> pour la date du calendrier
 - <time> pour l'heure
- Quelques exemples d'attributs :
 - @when pour la précision
 - @from @to pour un intervalle de temps certain
 - @notBefore @notAfter pour un intervalle de temps incertain
 - @calendar pour les calendriers pour <date>

La temporalité : exemples

```
<date when="1632-08-02">lundy  
second jour d'aoust</date>
```

Lundy second jour d'aoust

```
<time when="04:00:00">quatre  
heures du matin</time>
```

quatre heures du matin

```
<time from="23:00" to="24:00">  
entre ynze heures et  
minuict</time>
```

entre unze heures et minuict

```
<time notBefore="16:00"  
notAfter="17:00">quatre ou cinq  
heures de l'après-disnée</time>
```

quatre ou cinq heures de
l'apres-disnee

Les entités nommées

Extraction des chemins entre deux entités
nommées en vue de l'acquisition des patrons
de relations

Yayoi Nakamura-Delloye

ALPAGE, INRIA-Requencourt

Domaine de Voluceau Rocquencourt B.P.105 78153 Le Chesnay

http://hal.archives-ouvertes.fr/docs/00/51/14/81/PDF/P120_Poster62.pdf

- Définition :
 - « Tous les éléments du langage définis par référence : les noms propres au sens classique, les noms propres dans un sens élargi mais aussi les expressions de temps et de quantité » (Friburger, 2006).
 - Dans la phrase, « Martine Aubry dirige le PS » : deux entités nommées, une du type « individu » (Martine Aubry) et une du type « organisation » (PS).

Les entités nommés

- Plusieurs types : on en retient quatre
 - Personnes
 - Lieux
 - Les fonctions
 - Les organismes

Les entités nommés

- Problèmes :

- Lieux

- Orthographe et variante : « Bretagne »
 - Changement de noms de lieux

- Personnes

- Orthographe : « Lefebure » « Lefebvre »
 - Identité sociale

- Fonctions et organismes

- Limite imprécise entre ces quatre entités :
 - Exemple : le nom peut être détaché de sa fonction et de l'institution qui sont alors privilégiées par rapport au nom
 - Limite imprécise entre les titres et les fonctions

Les entités nommées et TEI

- **<rs>** referencing string (chaîne de caractères)

```
<rs type="person">Mme Nicole Dufournaud</rs>
```

```
<rs type="person">Charles de Gaulle</rs>
```

```
<rs type="airport">Charles de Gaulle</rs>
```

```
<rs type="office">Secrétaire d'Etat</rs>
```

```
<rs type="organization">ONU</rs>
```

```
<rs type="mythologie">Vénus</rs>
```

- **<name>**

```
<name type="person">Mme Nicole Dufournaud</name>
```

```
<name type="airport">Charles de Gaulle</name>
```

```
<name type="person">Charles de Gaulle</name>
```

```
<name type="mythologie">Vénus</name>
```

Les entités nommées et TEI : les noms de lieux

- Nom géographique :

```
<geogName type="airport">Aéroport  
<name>Charles de Gaulle</name></geogName>
```

- <geogFeat> *pour les noms communs*
(vallée, mont, rivière...)

- Nom de lieu :

```
<placeName>Roissy-en-France</placeName>
```

- <settlement> *pour les localités*
- <region> *pour les régions*

Les entités nommées et TEI : les noms de lieux

- `exercice_tp-6.xml`

`<geogName>` `@type`

`<country>` *pour les pays*

`<orgName>` *pour les organisations*

`@type="people"`

`<name>` *pour les noms de
personnes réels ou
mythiques*

Les entités nommées et TEI : les noms de personnes

<persName>

<surname> *nom de famille ou patronyme*

<forename> *prénom*

<genName> *generational name*

<roleName> *titre et avant-nom*

<addName> *épithète*

<nameLink> *particules nobiliaires*

Exemple d'indexation de nom de personne : Monsieur Ban Ki-moon, Secrétaire général de l'ONU

```
<persName>  
  <roleName type="avant-nom">Monsieur</roleName>  
  <forename>Ban</forename>  
  <surname>Ki-moon</surname>  
  <roleName type="fonction">Secrétaire  
général</roleName> de l'  
  <orgName type="organization">ONU</placeName>  
</persName>
```

Exemple d'indexation de nom de personne :

Haute dame dame Suzanne de Bourbon
dame de Rieux

```
<persName>  
  <addName type="epithete-honneur">haute dame</addName>  
  <roleName type="avant-nom">dame</roleName>  
  <forename>Suzanne</forename>  
  <nameLink>de</nameLink>  
  <surname type="maison">Bourbon</surname>  
  <roleName type="titre">dame</roleName>  
  <nameLink>de</nameLink>  
  <placeName type="terre">Rieux</placeName>  
</persName>
```

Exemple d'indexation de nom de personne : Catherine II la Grande impératrice de Russie

```
<persName>  
  <forename>Catherine</forename>  
  <genName>II</genName>  
  <addName type="epithete-honneur">la  
  Grande</addName>  
  <roleName>impératrice de Russie</roleName>  
</persName>
```

Base de données de noms et d'évènements

Les éléments et les attributs :

– <listPerson>

- <person> avec @xml:id
- <relation> avec @name @mutual

– <listEvent>

- <event> avec @xml:id @from @to @where
 - <label>
 - <head>

Liste d'événements

```
<p>
  <listEvent>
    <event xml:id="gen-01" when="1572-08-28">
      <head>Mariage</head>
      <label>Sacrement</label>
    </event>
  </listEvent>
  <lb/>Le 28 d'aoust 1572 nous nous sommes mariés moy
Jacques
  <lb/>de Bourgues et Françoise de Arnolet ma femme.
Dieu veille que <unclear>soict pour</unclear>
  <lb/>son saint service.
</p>
```

Base de données de noms et généalogie

```
<listPerson>
  <person xml.id="JDB">
    <persName>Jacques de Bourgues</persName>
  </person>
  <person xml.id="FDA">
    <persName>Françoise de Arnollet</persName>
  </person>
  <relation name="spouse" mutual="#JDB #FDA"/>
</listPerson>
```

Les noms de personnes

Exercice_tp-7.xml

- Utiliser les principaux éléments et attributs suivants :
 - <persName>
 - <forename> et <surname>
 - <genName>
 - <addName>
 - <nameLink>
 - <roleName>
 - <geogName>
 - <date>
 - <time>
 - <name>
 - @when et @type

Index et base de données

Index de personnes avec régularisation:

- Les éléments : `<index> <term>`
- L'attribut : `@indexName`

Exemple :

```
Le 10 octobre 1577 nasquit mon  
fils Jaques Bourques
```

```
<index indexName="index-person">
```

```
<term>Jacques de Bourgues</term>
```

```
</index>
```

Exercice sur les différentes entités nommées

Exercice_tp-8.xml

- Utiliser les éléments suivants :
 - <persName>
 - <roleName>
 - <orgName>
 - <placeName>