Formation à l'utilisation de TEI pour l'édition critique de manuscrits (École nationale des chartes, mars 2011) — Le travail de l'éditeur, suite : les notes, l'indexation

Florence Clavaud (École nationale des chartes)

1. Introduction

Après avoir établi le texte et fourni les premiers éléments de paratexte (une analyse, et la description du manuscrit), l'éditeur va enrichir le paratexte en ajoutant au fichier des notes (commentaires explicatifs, précisions sur le contexte...). Puis, si c'est utile pour le projet d'édition, il balisera diverses informations pour en préparer l'indexation; certains des mots ou expressions indexés dans le corps de la transcription pourront être liés à des notices décrivant les "objets" indexés (personnes, lieux, organismes, événements...).

2. Les notes

On va parler ici des notes génériques, qui sont à distinguer des notes d'apparat critique, qu'on parle d'une édition critique imprimée ou d'une édition structurée en TEI.

Habituellement, dans une édition imprimée, le texte des notes figure après la transcription, et un appel de note est posé dans le texte à l'endroit auquel la note est associée. Dans un fichier TEI, il y a deux manières d'encoder des notes :

- on peut encoder les notes directement à l'endroit où elles viennent expliciter ou contextualiser le texte établi
- on peut les encoder hors du texte, et poser des liens entre le texte et la note

2.1. Encoder les notes dans le texte

Un exemple suffira pour comprendre comment on procède (extrait simplifié de la première version de l'édition du chapitre de Beaurain du Cartulaire blanc de Saint-Denis, acte 5):

Cette solution est la plus simple et la plus rapide. Elle présente cependant quelques inconvénients, parmi lesquels :

- dans le cas où la même note devrait servir à plusieurs endroits, il faut en répéter le texte
- le texte de la note interrompt le texte transcrit, ce qui rend la relecture difficile par une personne (sauf si bien sûr on dispose de moyens appropriés pour cette relecture)

2.2. Encoder les notes hors du texte

Pour travailler selon cette méthode, il faut :

- choisir un emplacement hors du texte pour l'encodage des notes. Cet emplacement peut par exemple être l'élément <back> qui suit l'élément <body>; dans <back>, on va utiliser <div>
- encoder les notes à cet endroit, en utilisant pour chaque note un élément <note>.
 Puis affecter un identifiant (à l'aide de l'attribut @xml:id) à l'élément <note>.
- dans le texte, poser un élément <ref>, avec un attribut @target qui va pointer vers la note concernée (la valeur de @target est une chaîne de caractères constituée du dièse # et de l'identifiant de la note cible).

Exemple (extrait du dernier état de l'édition du chapitre de Beaurain, même acte 5):

```
<!-- dans la transcription -->
        Universitati vestre notum facimus quod nos partem nostram foreste desuper
               <foreign xml:lang="fro"</pre>
>Ivete</foreign>
           <ref type="note"
target="#beaurain-acte5-n-1"
/> et partem nostram
          de nemore [...]
        <!-- plus loin, dans l'élément back -->
        <div type="notes"
            <note xml:id="beaurain-acte5-n-1"</pre>
>Le nom se rapporte sans doute non au
                cours d'eau, mais au hameau d'Yvette, au bord de la rivière
               homonyme, sur la commune de Lévis-Saint-Nom (Yvelines, cant.
               Chevreuse).</note>
            <!-- autres notes -->
        </div>
```

Cette méthode n'a pas les inconvénients de la première méthode, par contre elle est un peu plus compliquée à mettre en œuvre.

On peut noter qu'avec cette méthode, rien n'empêche de préparer la création de un, deux, trois (ou plus, en fait autant qu'on veut) couches d'annotation.

3. Indexation d'entités nommées ou d'autres concepts

Toute expression contenue dans un texte encodé en TEI, qu'il s'agisse d'une entité nommée, d'un nom commun, d'une date, d'une catégorie juridique, ou d'un autre segment d'information, peut retenir l'attention de l'éditeur scientifique. Il décidera alors de la baliser au moyen d'un élément approprié.

Ce travail permettra, dans la page Web (ou la page imprimable) produite à partir du fichier XML, de formater de manière spécifique ces termes balisés. Il permettra surtout à un programme informatique de générer automatiquement un index de ces expressions. Dans certains cas, cet index pourra être rendu directement consultable (comme un index imprimé), ou encore il sera seulement utilisé à la recherche (une requête lancée depuis un formulaire retournera une liste des textes indexés avec la valeur saisie dans le formulaire).

Voici une petite liste des principaux éléments disponibles pour faire un tel travail : <name> (pour tout nom propre), ou encore (éléments spécialisés) <persName>, <geogName>, <placeName>, <orgName>; <rs>,<term>, <date>. La plupart de ces éléments contiennent des sous-éléments spécifiques(relatifs notamment aux segments qui les composent) ; à ce sujet voir la suite de cette présentation. Par ailleurs <persName>

peut contenir <placeName> ou <geogName>>, <term> et <rs> peuvent inclure des combinaisons de divers éléments.

Bien sûr il pourra être utile, voire nécessaire, surtout dans le cas où le concept à baliser est inscrit dans le texte édité, donc sous une forme peut-être non normalisée, d'établir un lien depuis l'élément utilisé pour baliser le concept vers une liste d'autorité, vers un glossaire ou même vers une notice d'autorité. Ces liens seront établis au moyen d'attributs (@ref et @key). De tels référentiels pourront être localisés, soit dans le même fichier TEI, soit dans un autre fichier TEI, soit dans une ressource distante (application Web, base de données locale). En ce qui concerne la description des entités nommées (la production de notices d'autorité en TEI), voir les sections ci-après.

Extrait de l'édition (en cours) d'un corpus de chartes médiévales (indexation, hors du texte transcrit, de la catégorie juridique d'acte):

Extrait simplifié de l'édition des minutes du notaire P. Christofle à Orléans en 1437 (indexation dans le texte de la minute):

```
Jehan Mignon et Belon sa femme, bourgois d'Orleans, ladicte femme
           deuement auctorisee etc., baillent a rente annuelle et perpetuelle pour
           touz jours a Mahiet Gauchier, paintre demourant a Orleans, qui a pris
           d'eulx pour lui, ses hoirs etc. ung jardin ouquel a ung poy de vignes et
           arbres, mouvens du propre heritaige de ladicte Belon, assis et seant
           leez l'eglise de <placeName xml:id="minute-0001-1-002"
ref="#1-0235"
                Saint Euvertre</placeName> d'Orleans en la rue appellee la
                <placeName xml:id="minute-0001-1-003"</pre>
ref="#1-0079"
>rue du Cimetire
               aux Juifs</placeName> es foursbours d'Orleans [...]
        <!-- plus loin, hors de la transcription -->
        tPlace>
           <!-- autres notices -->
            <place xml:id="1-0235"</pre>
                <placeName>Saint-Euverte</placeName>
                <!-- autres éléments -->
            </place>
            <!-- etc. -->
        </listPlace>
```

4. Encodage des informations relatives aux personnes

Les Guidelines TEI fournissent des moyens étendus pour l'encodage de telles informations. Au départ ces dispositifs ont été conçus dans l'idée de satisfaire les besoins des prosopographes, mais n'importe quel projet TEI peut en avoir besoin.

De telles informations incluent:

- les caractéristiques physiques telles que couleur des cheveux et des yeux,
- les caractéristiques culturelles telles que le statut social et la religion,
- des informations sur les activités et l'éducation, et les événements survenus dans la vie des personnes, tels que naissance, mariage ou emploi.

5. Principes de base

Les informations sur les personnes, à la base, sont regroupées au sein d'un élément <person>. Elles comprennent tout d'abord un élément spécialisé pour le nom de la personne, <persName>, et une série de déclarations relatives à :

des les caractéristiques ou qualités d'un individu, qu'elles soient physiques, traits comme le sexe, la couleur des yeux, ou culturelles, comme l'appartenance ethnique ou la religion. Certains de ces traits sont susceptibles d'évoluer plus que d'autres au fil de la vie de l'individu. Cependant aucun, en

principe, n'est immuable.

des ceci inclut, par exemple, le statut marital, le lieu de résidence, la position états

ou l'activité. De tels états ont normalement une durée de vie définie, i.e.

un début et une fin.

des les événements de la vie d'une personne tels que la naissance, le mariage ou événements l'emploi. De tels événements seront normalement associés avec une date

spécifique ou avec une fourcherre de dates assez étroite.

Toutes ces informations sont susceptibles d'être documentés et reliés à d'autres informations de même nature ou d'autre nature.

6. Traits, états et événements

Il existe trois classes d'éléments, en cohérence avec la classification tripartite présentée cidessus. Chacune de ces classes contient un petit nombre d'éléments spécifiques pour les catégories d'information les plus courantes, et un élément plus générique pour les autres catégories d'informations définies par l'encodeur.

Tous les éléments de ces trois classes appartiennent à la classe d'attributs datable, ce qui signifie qu'ils peuvent avoir les attributs

- @when pour une date unique,
- @notBefore et @notAfter, pour indiquer la date la plus ancienne et la date la plus récente possibles (le terminus ad quem et terminus ante quem) pour un trait, un état ou un événement.

7. Les traits

Les éléments spécifiques suivants sont disponibles pour encoder des traits :

- <langKnowledge>,
- <faith>,
- <nationality>,
- <age>,

- <socecStatus> ("statut socio-économique") et
- <sex>

Dans <langKnowledge> on peut utiliser soit des paragraphes soit des éléments <langKnown> ; ces deux sous-éléments peuvent avoir un attribut @tag (ou @tags), qui donne le code normalisé de la langue. L'élément <langKnown> possède aussi un attribut @level pour préciser le niveau de compétence de la personne dans la langue concernée.

On peut écrire:

L'élément <sex> possède un attribut **@value** qui doit contient un code choisi dans la liste de valeurs donnée par la norme ISO 5218:1977, i.e. 1 pour masculin, 2 pour féminin, 9 pour non applicable et 0 pour inconnu.

L'élément <trait> possède un attribut @type et peut contenir un élément <label>, qui peut être utilisé pour fournir la dénomination (lisible par l'humain) du trait concerné, suivi par un élément <desc>; alternativement, la description du trait peut être donnée au moyen d'un ou plusieurs éléments .

8. Les états

La classe d'éléments utilisable pour les états contient les éléments :

- <persName>,
- <occupation>,
- < <residence>,
- <affiliation>,
- <education> et

 <floruit>, pour spécifier pendant quelle période une personne donnée a été active (à notre connaissance), dans le cas où les dates exactes de naissance et de décès sont inconnues.

```
<floruit notBefore="1766"
notAfter="1790"
>fl. 1766-1790/floruit>
```

L'élément générique <state> a le même modèle de contenu que <trait>. L'exemple ci-après décrit le premier emploi occupé par le pasteur et poète islandais Jón Oddsson Hjaltalín:

```
<state type="office"</pre>
from="1777-04-07"
to="1780-07-12"
xml:lang="en"
        Jón's first living, which he apparently accepted
           rather reluctantly, was at
            <name type="place"</pre>
>Háls í Hamarsfirði</name>,
            <name type="place"</pre>
>Múlasýsla</name>, to which
            he was presented on 7 April 1777. He was ordained
            the following month and spent three
            years at Háls, but was never happy there,
            due largely to the general penury
            in which he was forced to live. In June of 1780
            the bishop recommended that Jón should
            <q xml:lang="da"
>promoveres til andet
            bedre kald, end det hand hidindtil har havt</q>,
            and on 12 July it was agreed that he should
            exchange livings with <name type="person"
key="#ThorJon"
            sr. Þórður Jónsson</name> at
            <name type="place"</pre>
>Kálfafell á Síðu</name>,
            <name type="place"</pre>
>Skaftafellssýsla</name>.
            <bibl>ÞÍ, Stms I.15, p. 733.</bibl>
            <bibl>ÞÍ, Stms I.17, p. 102.</bibl>
        </state>
```

9. Les noms de personne

L'élément <persName> est répétable et peut, comme tous les éléments TEI, avoir un attribut @xml:lang pour indiquer la langue du contenu de l'élément, ce qui permet de donner le nom d'une personne en plusieurs langues:

```
</person>
```

De nombreux sous-éléments sont disponibles dans persName> pour encoder les différentes parties d'un nom de personne. Voici un exemple d'utilisation de certains de ces éléments (extrait de l'édition des minutes notariales de P. Christofle, notaire à Orléans, pour l'année 1437) :

Voici un autre exemple, dans lequel les attributs @type et <sort> sont également présents.

```
<persName xml:lang="bg"</pre>
        <forename type="first"</pre>
 sort="2"
>######</forename>
        <forename type="middle"</pre>
 sort="3"
>#########</forename>
        <forename type="patronym"</pre>
 sort="4"
>#######</forename>
        <surname sort="1"</pre>
>#######</surname>
        <addName type="nick"
 xml:lang="bg"
>####</addName>
        <addName type="nick"
xml:lang="en"
>Alex</addName>
    </persName>
```

10. Les éléments relatifs aux événements

La classe des éléments relatifs aux événements contient des éléments spécifiques pour la naissance (<birth>) et le décès (<death>), et un élément générique <event>. Ce dernier a le même modèle de contenu que <state> et <trait> et peut être utilisé pour décrire tout événement de la vie d'un individu.

```
R. W. Dixon</name> with <name type="person"
ref="#CF"
            Charles Faulkner</name> as the best man.
            The bride was given away by her
            father, <name type="person"
ref="#RB"
>Robert Burden</name>.
           According to
            the account that <name type="person"
ref="#EBJ"
>Burne-Jones</name> gave
               <name type="person"</pre>
ref="#JWM"
>Mackail</name>
            <q>M. said to Dixon beforehand <q>Mind you don't call her Mary</q>
            but he did</q>. The entry in the Register reads:
            <q>William Morris, 25,
                Bachelor Gentleman, 13 George Street, son of William Morris decd.
                Gentleman. Jane Burden, minor, spinster, 65 Holywell Street, d. of
                Robert Burden, Groom.</q> The witnesses were Jane's parents and
            Faulkner. None of Morris's family attended the ceremony. Morris
            presented Jane with a plain gold ring bearing the London hallmark for
            1858. She gave her husband a double-handled antique silver cup.</desc>
        <bibl>J. W. Mackail, <title>The Life of William Morris</title>, 1899.</bibl>
    </event>
```

11. Utilisation de l'attribut @ref et de l'élément <relation>

Dans notre exemple l'attribut **@ref** posé sur chacun des éléments <name> permet de relier cet élément à un élément <person>, encodé ailleurs dans le fichier, qui décrit la personne concernée. Ci-dessous on voit qu'on peut aussi, grâce à l'élément <relation> et à ses différents attributs, déclarer une relation spécifique entre plusieurs personnes en faisant également référence aux mêmes éléments <person>.

12. Affectation de dates à ces éléments

Comme déjà indiqué, tous ces éléments, qu'ils soient spécifiques ou génériques, sont membres de la classe d'attributs *datable*, ce qui veut dire que leur contenu peut être daté, comme dans l'exemple ci-après, où l'on lit que la personne d'abord appelée David Jones a changé de nom en 1966 pour s'appeler David Bowie:

```
<person xml:id="DB"
```

13. Déclaration du degré de certitude et de la responsabilité

Tous les éléments génériques que nous avons vus sont aussi membres de la classe *editLike*, qui, comme son nom l'indique, a originellement été conçue pour fournir des attributs "décrivant la nature de l'intervention éditoriale ou de l'interprétation". Cette classe fournit les attributs @cert, pour indiquer le niveau de certitude, @resp pour indiquer le nom de l'entité responsable de l'information, et @evidence, pour préciser la nature de la preuve utilisée. Il est ainsi possible, dans le cas de sources multiples et non cohérentes, de fournir plus d'un point de vur sur ce qui est arrivé, comme dans l'exemple ci-après, où un lieu et une date de naissance différents ont été avancés pour le même individu par deux chercheurs, l'un étant considéré comme plus fiable que l'autre:

L'attribut @resp peut aussi être utilisé pour distinguer des objets qui sont explicitement mentionnés dans la source de ceux qui, même s'ils n'y sont pas explicitement mentionnés, peuvent être déduits des informations disponibles, comme dans l'exemple ci-après, tiré d'un ensemble de données réunies par Sebastian Rahtz sur les personnes enterrées au cimetière protestant de Rome, sur la base des informations trouvées sur leurs tombes :

14. Un autre exemple de notice descriptive de personne

Notice simplifiée tirée d'un fichier produit dans le cadre d'un projet de la BnF:

```
<person xml:id="f123456"</pre>
             <persName>
                <surname>Douceur</surname>
                 <forename>Louis</forename>
            </persName>
             <residence>
                 <placeName><settlement>Paris</settlement></placeName>
             </residence>
            <birth when="1700"</pre>
 cert="low"
>?</birth>
            <death when="1769-02-17"</pre>
>1769</death>
            <occupation>Relieur-libraire, créateur des reliures à décor de dentelle
                et de mosaïque.</occupation>
            <floruit notBefore="1721"</pre>
notAfter="1766"
>En activité de 1721 à
                1766.</floruit>
        </person>
```

15. Encodage des informations sur les lieux

Les mêmes possibilités existent pour le balisage des lieux géographiques, ce qui permet de normaliser ou d'uniformiser les références aux lieux, en les décrivant puis en associant ces descriptions avec une collection de textes balisés, par exemple des documents historiques balisés.

Tout comme pour les personnes, les informations sur les lieux géographiques consistent en une série de déclarations concernant:

- tels que le climat, la localisation, etc..
- tels que la population, l'aire géographique, le statut administratif.
- comme les circonstances d'une fondation, de catastrophes naturelles ou causées par l'homme.

Et comme pour les déclarations relatives aux personnes, chacune de celles-ci doit pouvoir être documentée, être inscrite dans un cadre temporel et être reliée à d'autres de même nature ou d'autre nature.

15.1. L'élément <place>

L'élément <place> possède un attribut @type qui permet de caractériser brièvement le lieu, par exemlpe de dire qu'il s'agit d'une ville, d'une rue, d'un immeuble, d'un objet naturel tel qu'un lac ou une montagne, ou un lieu célèbre comme un champ de bataille. L'élément <place> contient:

- un ou plusieurs éléments <placeName>, chacun donnant un nom sous lequel le lieu est ou a été connu (les éléments TEI <country>, <bloc>, <geogName>,
 <region> et <settlement> sont des spécialisations de l'élément générique
 <placeName>, et peuvent être utilisés à la place de <placeName> ou dans
 <placeName>).
- un ou plusieurs éléments <location>, précisant la localisation physique du lieu, sous la forme, par exemple, d'un jeu de coordonnées géographiques, d'une adresse, ou de la mention de l'appartenance à une entité géo-politique.

15.2. Où saisir un élément <place>

Les éléments <place> sont des enfants de listPlace>, qui, comme listPerson>, est le plus souvent placé dans l'en-tête TEI (dans <sourceDesc>), mais qui peut aussi apparaître à tout point du fichier où une liste ordinaore (un élément list>peut aussi apparaître.

```
<place type="lake"</pre>
xml:id="Skadar"
        <placeName xml:lang="sr-Latn"</pre>
>Skadarsko jezero</placeName>
        <placeName xml:lang="sr-Cyrl"</pre>
>######## ####</placeName>
        <placeName xml:lang="sq"</pre>
>Liqeni i Shkodrës</placeName>
        <placeName xml:lang="en"</pre>
>Skadar Lake</placeName>
        <location>
             <geo>42.166389 -19.325833
        </location>
        <location>
             <desc>The lake is on the border between
                 <name type="place"</pre>
 kev="ME"
>Montenegro</name>
                 and <name type="place"
 key="AL"
>Albania</name>,
                 with about two-thirds of it lying in the
                 territory of the former, one-third in the latter.</desc>
         </location>
    </place>
    <place xml:id="1-0093"</pre>
        <placeName>Corroy</placeName>
        <location>
             <district type="departement"</pre>
>Loiret</district>
            <district type="canton"</pre>
>Saint-Jean-de-la-Ruelle</district>
             <settlement type="commune"</pre>
```

```
>Ormes</settlement>
        </location>
    </place>
    <place xml:id="1-0196"</pre>
         <placeName>Orléans</placeName>
         <location>
             <district type="departement"</pre>
>Loiret</district>
        </location>
        <place xml:id="d3e4180"</pre>
             <placeName>
                 <term>châtelet</term>
             </placeName>
         </place>
         <place xml:id="d3e4182"</pre>
             <placeName>
                 <term>duché</term>
             </placeName>
        </place>
         <place xml:id="d3e4184"</pre>
             <placeName>
                 <term>église</term>
             </placeName>
         </place>
    </place>
```

15.3. Les traits

Il existe, en plus d'un élément générique <trait>, des éléments spécialisés pour la population (<population>), le climat (<climate>) et le terrain (<terrain>).

```
<
```

15.4. Les états

Les états sont traités grâcé à l'élément générique <state>.

</state>

15.5. Les événements

Il n'existe pas d'éléments spécialisés pour les événements ; on utilisera l'élément générique <event> qui peut être doté d'un attribut @type:

```
<event type="political"</pre>
     when="2006-06-03"
   <head>Independence</head>
   On <date when="2006-05-21"</p>
    >21 May 2006</date>
 a referendum was held in Montenegro on whether
  the state union with <name type="place"
     kev="RS"
     >Serbia</name> should be dissolved.
  A total of 419,240
  votes were cast, of which 230,661
  were for independence.
   Montenegro's referendum commission officially confirmed the results
  of the referendum on <date when="2006-05-31"
    >31 Mav</date>,
  verifying that 55.5% of the population had voted in favour of
  independence, narrowly surpassing the threshold requirement of 55%
  set by the <name type="org"
     key="EU"
    >European Union</name>. The
  Montenegrin <name type="org"
     key="USRCG"
    >Parliament</name> made a
  formal declaration of independence on
  <date when="2006-06-03"</pre>
    >Saturday, 3 June</date>.
</event.>
```

Les valeurs des attributs **@key** des éléments <name> établissent un lien vers d'autres éléments <place>, vers des éléments <org> ou <person>.

16. Listes d'événements

Les événements, comme les personnes, les lieux - et les organismes -, peuvent être considérés comme des objets à décrire pour eux-mêmes. Dans ce cas, on utilisera l'élément listEvent> pour englober une série d'éléments <event>.

17. Encodage des informations sur les organismes

TODO: listOrg, org, et ses sous-éléments