

# Réflexions sur l'utilisation de la TEI pour coder les sources diplomatiques

## à partir de l'exemple du Cartulaire blanc de l'abbaye de Saint-Denis

Gautier Poupeau, [gpoupeau@enc.sorbonne.fr](mailto:gpoupeau@enc.sorbonne.fr), Ecole nationale des chartes.

Depuis maintenant deux ans, l'Ecole des chartes a décidé de mettre en place un programme d'édition électronique dans le cadre de son site Web et de sa politique de publication. L'édition électronique présentait en particulier l'avantage de proposer les résultats d'une entreprise d'édition telle que le Cartulaire Blanc<sup>1</sup> qui compte plus de 2500 actes au fur et à mesure de sa réalisation par les élèves sous la direction d'Olivier Guyotjeannin.

Dès la conception du projet, il nous a semblé important de réfléchir à l'utilisation du XML avec, au départ, un projet de construction ex-nihilo d'une DTD spécifique aux sources diplomatiques. Dans un premier temps, nous avons voulu tester rapidement les possibilités de l'édition électronique et avons donc mis en place un site dynamique basée sur le langage PHP<sup>2</sup> et une base de données avec le SGBD MySQL<sup>3</sup>. Cette solution transitoire présentait l'avantage d'entamer la réflexion sur la modélisation de l'édition critique de sources diplomatiques, sans présager de sa possible transformation en DTD et, surtout, d'arriver rapidement à un résultat qui avait valeur de prototype. Au terme de la programmation, nous avons mis en ligne la première version du Cartulaire blanc<sup>4</sup>.

A partir de cette expérience, nous avons pu mesurer les avantages de l'édition électronique et les limites du système mis en place. Ainsi, nous nous sommes aperçu que les possibilités d'une base de données étaient limitées par rapport aux besoins des chercheurs. Le niveau de granularité d'une base de données peut approcher celui du XML au prix de la mise en place d'une modélisation qui devient rapidement trop complexe et donc moins gérable. L'utilisation du XML et d'une DTD devenait donc indispensable pour permettre un balisage plus fin de l'information et un niveau de profondeur plus important que ceux offerts par une base de données contenant des informations en HTML. De plus, la base de données présente le désavantage de devoir mettre en place une modélisation spécifique pour chaque édition.

---

<sup>1</sup> Pour plus de renseignements sur le Cartulaire blanc, Guyotjeannin Olivier et Gautier Poupeau, « Le projet d'édition électronique du Cartulaire blanc de l'abbaye de Saint-Denis et les projets d'édition électronique de l'École des chartes », *Le médiéviste et l'ordinateur*, n°42, printemps 2003, la diplomatique, éd. IRHT, Paris, 2003, [en ligne], [http://lemo.irht.cnrs.fr/42/mo42\\_12.htm](http://lemo.irht.cnrs.fr/42/mo42_12.htm).

<sup>2</sup> Site officiel de PHP : <http://www.php.net/>

<sup>3</sup> Site officiel de Mysql : <http://www.mysql.com/>

<sup>4</sup> Olivier Guyotjeannin (dir.), *Le Cartulaire blanc*, Paris, 2001, [en ligne], <http://www.enc.sorbonne.fr/cartulaireblanc/>

Plutôt que de créer une DTD ex-nihilo, il nous a paru plus intéressant d'étudier la possibilité d'utiliser une DTD préexistante. Nous nous sommes rapidement tournés vers la TEI<sup>5</sup>. L'étude que nous en avons faite a montré qu'elle semblait correspondre à nos besoins pour les sources diplomatiques<sup>6</sup>. Surtout, son utilisation présentait pour nous l'avantage de n'utiliser qu'une seule DTD pour tous les types de sources que nous éditons, nos projets ne se limitant pas aux seules sources diplomatiques<sup>7</sup>. Nous pouvions à long terme envisager la mise en place de systèmes simples d'interrogations croisées entre différents types de sources.

Nous avons pris la décision d'utiliser la TEI en juillet 2003. Depuis, nous sommes dans une phase de test qui nous a conduits à l'élaboration de deux sites Web non encore rendus publics à l'heure d'aujourd'hui : l'édition de l'Obituaire du Saint-Mont par Marie-José Gasse-Grandjean et la deuxième version du Cartulaire blanc intégralement en XML. Nous ne prétendons donc pas avoir une expérience encore très importante de la TEI ; pour autant, nous nous efforcerons de montrer dans cet article comment nous utilisons cette DTD pour coder le Cartulaire blanc de l'abbaye de Saint-Denis, en vous faisant part de notre réflexion qui a accompagné le choix des balises, et ceci en fonction de notre problématique générale : proposer une édition critique au format électronique.

Nous avons adopté un parti-pris qu'il nous semble intéressant d'expliquer avant de décrire notre utilisation de la TEI. Avant de commencer le balisage d'une édition, nous commençons par déterminer en compagnie de l'éditeur scientifique le niveau de balisage qu'il souhaite adopter et les différentes balises qu'il souhaite utiliser, en dehors d'un jeu de balises par défaut qui reprend les pratiques de la publication papier. Pour cela, nous avons édicté une règle : toutes les informations balisées doivent avoir une utilité : soit pour mettre en valeur graphiquement un élément par rapport au reste du texte, soit pour refaire sortir cet élément dans le cadre d'une recherche précise ou dans des index. Ce choix s'explique, en partie, par le temps que prend le balisage fin d'un texte ; ainsi, le balisage reflète les besoins et les problématiques de l'éditeur scientifique, sans préjudice du fait qu'un autre utilisateur pourra reprendre notre fichier XML pour le baliser en vue d'une autre problématique ou d'un autre intérêt.

---

<sup>5</sup> Site du consortium TEI : <http://www.tei-c.org>

<sup>6</sup> Cet étude a été menée avec l'aide de Jean-Daniel Fekete, chercheur à l'INRIA et réalisateur du site, les lettres de rémission de la chancellerie de Bretagne en TEI, <http://nicole.dufournaud.net/remission/>.

<sup>7</sup> Pour plus de renseignements sur les projets de recherche de l'Ecole et les futures éditions électroniques : <http://www.enc.sorbonne.fr/enseignement/recherche.htm>.

# I- Problème de structuration générale

## A- Au niveau de l'édition complète

Le but des éditions électroniques de l'Ecole des chartes est avant tout de proposer aux chercheurs une édition critique de sources, reprenant les caractéristiques de la publication papier tout en proposant des fonctionnalités propres à l'édition électronique. Nous concevons donc notre balisage XML dans cette optique, et non dans celle d'une étude diplomatique, paléographique ou lexicographique. Ponctuellement, l'éditeur scientifique peut nous demander de baliser des informations complémentaires en rapport avec une problématique qui l'intéresse, mais ce n'est pas le cas général.

Ainsi, l'acte n'est pas l'unité de base pour construire le balisage, mais l'édition critique dans son ensemble. Nous avons donc un seul fichier XML par édition critique. Cette caractéristique nous a obligés à réfléchir aux différentes granularités d'une édition critique de sources diplomatiques en vue de les baliser.

Nous avons pu dégager les parties suivantes :

- Partie introductive à l'édition : présentation, analyse de l'édition ;
- Un corpus d'actes. Un acte peut être considéré comme un texte à part entière et donc faire l'objet d'un seul fichier XML. Mais, dans le cadre d'une édition, l'acte est appréhendé comme la partie d'un corpus ;
- Les annexes : glossaire, index, documents annexes.

Nous avons pu, assez facilement, adapter ce schéma avec la TEI. Un fichier balisé avec cette DTD comprend deux parties contenues dans une balise racine <TEI.2></TEI.2> :

- Un en-tête ou header contenant toutes les informations sur l'édition : <teiHeader></teiHeader> ;
- Et le corps de l'édition <text></text>.

Deux méthodes permettent de coder un corpus en TEI :

- la balise <teiCorpus.2></teiCorpus.2> à la racine et autant de <TEI.2></TEI.2> que de textes. Cette méthode est plutôt employée pour des textes hétérogènes comme une anthologie

- la balise `<group></group>` à l'intérieur de la balise `<text></text>` qui permet ensuite de répéter l'élément `<text></text>` autant de fois qu'il y a de textes. Cette solution est plutôt conseillée dans le cas de textes homogènes.

Notre choix s'est tourné tout naturellement vers la deuxième solution. La construction générale de notre fichier TEI se présente donc ainsi :

```

<TEI.2> La racine du fichier
  <teiHeader> L'en-tête du document contenant toutes les
  </teiHeader> métadonnées de l'édition.
  <text>
    <front> Le <front> général contenant les parties
    </front> introductives de l'édition.
    <group>
      <text>
        <body> L'acte dont nous détaillerons plus loin
        </body> la structuration.
      </text>
    </group>
    <back> Le <back> général contenant toutes les annexes de
    </back> l'édition.
  </text>
</TEI.2>

```

Il est à signaler que dans le cas particulier du Cartulaire blanc qui est découpé en chapitres, nous utilisons la balise `<teiCorpus.2></teiCorpus.2>` pour l'ensemble de l'édition et, au niveau du chapitre, la balise `<TEI.2></TEI.2>`. Il faut ajouter que la description exposée précédemment convient parfaitement à l'édition d'actes, mais que la TEI est assez souple pour pouvoir coder d'autres types de sources diplomatiques qui n'ont pas la même granularité qu'un recueil d'actes.

## ***B- Au niveau de l'acte***

Comme nous l'avons expliqué précédemment, chaque acte constitue un texte en soi, l'édition critique étant le rassemblement de ces différents textes au sein d'un corpus défini par

l'éditeur scientifique en fonction de son intérêt et des sources à sa disposition (cartulaires, chartiers laïques ou ecclésiastiques, recueil d'actes...).

Nous retrouvons donc au niveau de l'acte un schéma équivalent à celui mis en avant pour l'édition :

- Une partie introductive à l'acte, contenant les **métadonnées** et l'analyse ;
- Le texte de l'acte ;
- Les annexes concernant exclusivement l'acte.

Il est à signaler que pour toutes les informations ne disposant pas de balises spécifiques en TEI, nous utilisons l'élément désignant la division d'un texte avec la précision du type d'informations contenues, exemple : `<div type= « regeste »></div>` pour le regeste.

Chaque élément `<text>` est composé de trois parties :

- L'élément `<front></front>` rassemblant les métadonnées de l'acte et les différentes analyses.
- L'élément `<body></body>` contenant le tableau de la tradition et le texte de l'acte  
Nous avons fait le choix de ne pas coder les différentes parties du discours diplomatique. Ce choix s'explique par le fait que nous n'avons pas une approche strictement diplomatique, mais celle d'une édition critique. Pour autant, cette possibilité reste accessible en utilisant la balise `<div>` avec la précision du nom de la partie dans l'attribut `type`, par exemple : `<div type= « corroboratio »></div>`.
- L'élément `<back></back>` qui contient les annexes de l'acte : documents complémentaires, bibliographies sur l'acte.

## II- Codage des informations relatives à l'édition critique de sources diplomatiques

### A- Le tableau de la tradition

L'élaboration du tableau de la tradition de l'acte par l'éditeur scientifique est fondamentale dans la mise au point d'une édition critique d'un acte, l'analyse de ce tableau étant une partie intégrante de toute étude diplomatique sérieuse. C'est pourquoi l'existence de balises spécifiques en TEI pour coder le tableau de la tradition a été déterminant dans notre choix pour cette DTD. De plus, l'intérêt du codage est de pouvoir faire des recherches précises sur les différents éléments de l'édition ; dans cette perspective, baliser précisément le tableau de la tradition permet d'offrir des interrogations croisées et de constituer des statistiques et des graphiques sur l'édition comme sur l'état de la tradition.

Le tableau de la tradition est codé avec l'élément <witList>. Chaque témoin est codé avec la balise <witness>, le code correspondant à la nature du témoin (suivant en cela les règles de la Commission internationale de diplomatique) étant indiqué avec l'attribut sigil. Exemple :

```
<witList>
  <witness sigil="A">Original sur parchemin</witness>
</witList>
```

Le balisage nous a obligés à réfléchir aux différentes informations disponibles dans un tableau de la tradition :

- Informations diplomatiques :
  - o Etat du document (original, copie, notice de cartulaire, fac-similé, traduction, regeste)
  - o Date de la copie <date>
  - o Source du témoin <witDetail type= « source »></witDetail>
- Informations codicologiques :
  - o Support de l'acte : <witDetail type= « support »></witDetail>
  - o Les mesures du parchemin : <measure>
  - o Présence et description d'un chirographe : <witDetail type= « chirographe »></witDetail>
  - o La description du scellement : <witDetail type= « scellement »></witDetail>

- La description des mentions dorsales : `<witDetail type= « mentions_dorsales »></witDetail>`
- Informations bibliographiques et/ou archivistiques : `<bibl></bibl>`
  - Identification `<title></title>`
  - Lieu de conservation `<placeName></placeName>`
  - Cote `<idno></idno>`
  - Précision du folio : `<biblScope type= « folio »></biblScope>`
  - Précision de la rubrique `<biblScope type= « rubrique »></biblScope>`
  - Précision du numéro dans le carton d'archive `<biblScope type= « numero »></biblScope>`

Cette liste tient compte de tous les types d'informations et leurs balises correspondantes, mais la finesse du codage du tableau de la tradition est laissée à l'appréciation de l'éditeur scientifique, comme pour le reste de l'édition.

## ***B- L'établissement du texte***

Nous n'aborderons dans cette partie que les informations que les éditeurs ont déjà l'habitude de traiter dans une édition traditionnelle. La question des informations qu'on peut exploiter grâce à un balisage du texte sera traitée plus loin.

Nous pouvons identifier deux types d'informations que l'éditeur peut être amené à indiquer pour l'établissement du texte :

- Les informations indiquées de manière systématique par l'éditeur suivant des règles énoncées ou rappelées en introduction et qui n'appellent donc aucun commentaire ;
- Toutes les informations qui relèvent de l'apparat critique, c'est à dire tous les commentaires ponctuels que l'éditeur est amené à faire sur l'établissement ou la présentation du texte.

### **1- Informations systématiques**

Les informations systématiques sont de différentes natures. Elles peuvent être d'ordre philologique, par exemple, les mots ou expressions exprimés dans une autre langue que la langue majoritaire du texte pour lesquels on utilise la balise `<foreign></foreign>` avec la

précision de la langue dans l'attribut lang. La langue est codée suivant la norme ISO 639-2<sup>8</sup>. Elles sont aussi d'ordre graphique :

- Les abréviations peuvent être codées de deux manières différentes : soit on balise l'abréviation <abbr></abbr> et la résolution est dans l'attribut expan, soit on balise le mot résolu <expan></expan> et l'abréviation est l'attribut abbr. Nous avons préféré la deuxième solution. La règle dans l'édition critique étant de résoudre les abréviations, il nous a paru plus logique de coder la résolution. On peut aussi donner le degré de certitude dans la résolution de l'abréviation avec l'attribut certain.
- Lorsqu'elle est précisée, l'indication du changement de lignes dans le texte édité est donnée sous la forme /numéro de ligne, on peut utiliser la balise vide <lb/> et la précision du numéro de ligne dans l'attribut n. Si la source éditée est présentée sous forme de colonnes, on peut baliser cette information avec <cb/>. Enfin, le changement de pages, dans l'édition d'un cartulaire par exemple, peut être précisé avec <pb/>.
- Variations graphiques dans le traitement des lettres : mots en capitales, caractères allongés : <hi></hi> et l'attribut rend pour indiquer le type de variations.
- Un espace laissé blanc sur le texte édité, pour lequel on utilise habituellement le signe diacritique (...) : <damage></damage>
- Les signes graphiques : croix, chrismon, monogramme... peuvent être codés avec la balise <figure/> en précisant le type de signe dans l'attribut n et éventuellement en renvoyant à l'image numérisée du signe dans l'attribut rend.<sup>9</sup>

## 2- L'apparat critique

Par sa nature, l'apparat critique est la partie de l'édition qui fait l'objet de la réflexion la plus importante pour l'adaptation au balisage et à l'édition électronique. En effet, il a pour vocation d'accueillir tous les commentaires de l'éditeur sur l'établissement du texte ou des remarques ponctuelles sur la présentation de l'acte. Il accueille donc des informations très hétérogènes et qui appellent de ce fait des codages différents. Le système de l'apparat critique correspond à une habitude de travail issue de la publication papier pour laquelle il était plus pratique de renvoyer ces informations en notes. Ainsi, le fait de baliser précisément ces informations nous oblige à dresser une typologie fine des différents types d'informations et de la façon de les traiter ; cela représente un travail d'homogénéisation qui n'est pas toujours fait dans les publications papier :

---

<sup>8</sup>ISO 639-2 Registration Authority - Library of Congress, <http://www.loc.gov/standards/iso639-2/>.

<sup>9</sup> Nous aborderons plus loin le couplage entre un texte balisé et l'image numérisée.

- Les abréviations non résolues : <abbr></abbr>
- Mot ou passage restauré par conjecture : <reg></reg>
- Lacune comblée d'après d'autres sources : <supplied></supplied>
- Lacune impossible à combler : <gap/>
- Passage supprimé par la suite : <del></del> avec en complément des explications dans une balise <witDetail></witDetail>
- Addition postérieure : <add></add> avec attribut place pour déterminer sa place ;
- Lecture ou graphie curieuse : <sic></sic>
- Variante entre différents états du texte : <app><lem> - Leçon choisie par l'éditeur scientifique pour l'affichage</lem><rdg wit=« le témoin »> - Lecture d'un témoin précis</rdg></app>

Pour tous les autres types d'informations renvoyées dans l'apparat critique, nous utilisons la balise <witDetail></witDetail> en précisant le type d'informations dans l'attribut type. De plus, cette balise peut aussi être utilisée pour compléter une information donnée par les balises décrites précédemment.

Au moment de l'affichage, nous rassemblons tous ces types de notes en bas de page sous la forme habituelle de l'apparat critique, mais on peut également les faire apparaître sous forme de bulle, voire les intégrer au texte. De fait, nous pourrions utiliser la balise <note></note> ou seulement la balise <witDetail></witDetail> pour coder l'apparat critique et considérer toutes ces informations comme une note de bas de page. Mais, il nous a semblé plus utile de décrire précisément l'information, pour permettre l'établissement de listes ou une étude précise des différents phénomènes repérés par l'éditeur scientifique.

## **C- Les notes**

Au moins deux types de notes de bas de page sont utilisés par les éditeurs scientifiques : les notes dites historiques (identifications, précisions chronologiques, renvoi à d'autres actes, identifications de citations...) et les notes d'apparat critique. Nous avons vu précédemment les choix que nous avons faits concernant l'apparat critique. En ce qui concernent les notes « historiques », dont l'appel de note est le plus souvent un nombre, nous utilisons la balise <note></note> qui contient le texte de la note. Cette balise se place à l'intérieur du texte à l'endroit de l'appel de notes. La précision de l'appel de notes est donnée avec l'attribut n.

Il peut arriver que les éditeurs utilisent d'autres types de notes, en particulier pour identifier des citations précises, comme les citations bibliques. Pour cela, nous pouvons évidemment utiliser la balise `<note></note>` ; mais nous préconisons plutôt dans ce cas précis l'utilisation des balises ad-hoc, comme pour l'apparat critique. Ainsi, les citations précises et identifiées sont codées avec la balise `<cit></cit>` encadrant la citation dans le texte avec la référence bibliographique dans la balise `<bibl></bibl>`. Exemple :

```
<cit>
  <quote>Citation</quote>
  <bibl>Référence bibliographique</bibl>
</cit>
```

Si la citation appelle un commentaire de la part de l'éditeur, par exemple en cas de citation approximative, il sera rejeté en note et la balise `<note></note>` sera donc utilisée.

### **III- Les nouvelles possibilités ouvertes par le balisage**

Comme nous l'avons déjà signalé pour l'apparat critique ou les notes de bas de page, le balisage du texte ne présage pas de son affichage à l'écran. Ainsi, avec le balisage XML, nous faisons une distinction claire entre le contenu et l'affichage. Pour restituer à l'écran ou sur d'autres médias, nous utilisons donc le langage XSL, défini par le W3C<sup>10</sup>. Il nous a semblé important d'apporter cette précision, avant d'aborder la dernière partie, concernant les nouvelles possibilités ouvertes par le balisage.

#### ***A- Les métadonnées de l'édition et de l'acte***

Le balisage du texte donne la possibilité d'ajouter des métadonnées précises sur l'édition et sur les différents actes. Ces métadonnées nous permettent de générer des tableaux récapitulatifs des actes, de proposer aux utilisateurs des formulaires d'interrogations croisées et de pouvoir générer des métadonnées au format Dublin Core<sup>11</sup> précises en vue du référencement par les moteurs de recherche.

Toutes les informations concernant l'édition dans son ensemble sont contenues dans l'entête du fichier : <teiHeader>. Elles se décomposent en quatre parties :

- Les informations bibliographiques (<fileDesc>) ;
  - o Titre ;
  - o Editeur scientifique ;
  - o Informations éditoriales ;
  - o Droits du document ;
  - o Description de la source du texte édité.
- Les informations sur le codage du texte (<encodingDesc>) expliquant les pratiques de balisage du texte ;
- Les aspects non-bibliographiques du texte (<profileDesc>) comme les langues utilisées (<langUsage>) ;
- Les différentes révisions du texte édité (<revisionDesc>).

---

<sup>10</sup> Recommandations de XSL : <http://www.w3.org/Style/XSL/>

<sup>11</sup> Norme de métadonnées établi par le Dublin Core Metadata initiative : <http://www.dublincore.org/>

Au niveau de l'acte, la TEI permet d'ajouter les informations suivantes :

- L'auteur de l'acte : <docAuthor> ;
- La date de temps et de lieu : <docDate> précision du type de date (actio, recognitio, ...) avec attribut n et précision d'une valeur de date sous la forme normalisée : YYYY-MM-DD avec l'attribut value, la date de lieu est codée avec la balise <geogName><sup>12</sup> ;
- La langue principale de l'acte : attribut lang dans la balise <text> et/ou dans le <div type= « acte »> ;
- Pour le cartulaire blanc, un résumé de l'acte, plus court que le regeste a été ajouté <div type= « regeste\_bref »></div> en vue de la génération d'une table des actes, le regeste étant parfois trop long pour cela.

## ***B- Identifications et index***

Certains éditeurs scientifiques proposent des identifications de personnes ou de lieux à la fin de chaque acte et non en notes de bas de page. Dans le contexte de l'édition électronique, on peut baliser les noms de lieux et les noms de personnes avec respectivement <geogName> et <persName> avec un attribut corresp qui renvoie à l'identification sise dans le <back> général sous forme de liste

<geogName corresp= « i1 »>Lieux</geogName> dans le texte

```
<back>
  <div type="identification">
    <list>
      <label> Lieu</label>
      <item id="i1">Identification</item>
    </list>
  </div>
</back>
```

---

<sup>12</sup> Voir en annexe, le système mis au point pour coder les dates approximatives.

La TEI intègre une balise vide pour permettre l'indexation. Son fonctionnement se rapproche du fonctionnement de l'indexation sous le traitement de texte Word. Ainsi, la TEI n'indexe pas un mot ou un groupe de mots mais un point dans le texte. Exemple :

Romarimont, en la grant rue<index index="nominum" level1="Remiremont (Romarimont)" level2="rues" level3="grand-rue de la Salnerie"/>

### **C- Informations graphiques et numérisation**

Le but d'une édition critique n'est pas de donner une copie à l'identique du texte édité, cette copie étant forcément infidèle par rapport à la source éditée, mais, plutôt de proposer aux lecteurs des objets d'études scientifiques. Ainsi, le relevé systématique de certaines informations sur la présentation graphique de l'acte peut être pertinent dans certains cas. Par exemple, il nous semble utile de relever systématiquement les abréviations dans le but de faire une étude de l'utilisation des abréviations, et non dans le but de proposer un calque du texte édité. On peut aussi citer le cas de la ponctuation originale qu'il est absurde de restituer dans le cas d'une édition critique mais qui peut être pertinent dans le cas d'une étude de l'utilisation de la ponctuation. La ponctuation ancienne peut ainsi être codée avec la balise <c type= « ponctuation »></c>. La TEI propose aussi un jeu de balises permettant de mettre en place une étude paléographique précise, en particulier les changements de mains : <handSchift/>. Enfin, la possibilité donnée par l'édition électronique d'associer l'image numérisée au texte balisé nous semble plus pertinente que de faire une édition imitative<sup>13</sup>.

### **D- Autres indications possibles**

N'utilisant pas ou pas encore les balises décrites ci-après, nous nous contenterons d'en dresser la liste, en vous renvoyant au Guidelines TEI<sup>14</sup> pour plus de précisions.

Ainsi, il est possible de noter des terminologies avec les étiquetages du type Glossaire détaillé dans le chapitre « Terminological Databases » du Guidelines<sup>15</sup>. Les balises décrites dans le chapitre « Simple analytic Mechanisms »<sup>16</sup> permettent de baliser des termes ou groupes de

---

<sup>13</sup> Pour plus de renseignements concernant ce sujet, voir l'article de Jean-Daniel Fekete, « Expérience de codage de document à intérêt graphique à l'aide de TEI », dans *Actes du congrès Eurotex 98*, Saint Malo, p. 131-142. mars 1998, [en ligne], <http://www.lri.fr/~fekete/ps/eurotex2.pdf>.

<sup>14</sup> TEI Consortium, *TEI P4, Guidelines for Electronic Text encoding and Interchange*, 2001, [en ligne], <http://www.tei-c.org/P4X/>.

<sup>15</sup> <http://www.tei-c.org/P4X/TE.html>

<sup>16</sup> <http://www.tei-c.org/P4X/AL.html>

termes comme support d'une analyse historique, linguistique ou statistique, en particulier la balise `<ana></ana>`. Enfin, le codage des éléments statistiques ou quantitatifs utiles pour l'édition d'un livre de comptes ou d'un pouillé par exemple est décrit dans le chapitre Feature structures<sup>17</sup>.

Pour coder toutes les informations décrites précédemment, il est évident que la TEI ne va pas aussi loin que des DTD spécialisés ou d'autres systèmes d'équitages. Pour autant, la TEI offre un certain nombre de balises génériques dont on peut définir la fonction grâce aux attributs, comme nous l'avons vu avec la balise `<witDetail></witDetail>`.

---

<sup>17</sup> <http://www.tei-c.org/P4X/FS.html>

La généricité de la TEI et sa richesse permettent donc de coder des informations diverses, relevant de problématiques variées, mais qui peuvent se rassembler au sein d'un même fichier : l'édition critique de sources, l'analyse diplomatique, l'analyse paléographique et codicologique, l'analyse statistique et lexicographique et l'analyse historique. De plus, cette DTD offre la possibilité de coder les informations en fonction de leur rôle intellectuel dans le texte, sans se limiter a priori par des questions d'affichage et de rendu graphique. Ainsi, des listes d'éléments et de nombreuses clés d'interrogations peuvent être proposées aux chercheurs consultant cette édition électronique<sup>18</sup>. L'intérêt pour l'Ecole nationale des chartes d'utiliser la TEI réside aussi dans le fait que cette DTD n'est pas spécifique aux sources diplomatiques et qu'elle permet aussi de coder tous types de sources et ainsi, d'envisager l'interrogation croisée de différentes sources. Par exemple, le *Cartulaire blanc* pourra être interrogé en même temps que les *Chroniques latines de Saint-Denis*, qui sont actuellement en préparation sous la direction de Pascale Bourgain. Les possibilités de codages hétérogènes avec une seule DTD permettent aussi d'envisager des systèmes d'annotations collaboratifs, dans lesquels des chercheurs pourront récupérer un texte balisé en TEI, en vue d'y intégrer les balises qui les intéressent et de reverser ensuite leur codage pour que d'autres profitent de leur travail. Enfin, il faut souligner que le consortium TEI est actuellement en train de préparer une nouvelle version de la DTD, ou plutôt schéma XML<sup>19</sup>, qui marquera des changements importants concernant la transcription et une réflexion importante concernant les manuscrits, avec l'intégration de la DTD Master et la constitution d'un groupe de travail sur les manuscrits<sup>20</sup> dont le travail profitera sûrement aux sources diplomatiques.

---

<sup>18</sup> L'ensemble des éditions électroniques de l'Ecole des chartes sont accessibles sur le Web librement et gratuitement : <http://elec.enc.sorbonne.fr> ou <http://theleme.enc.sorbonne.fr>.

<sup>19</sup> Pour plus de renseignements, <http://www.tei-c.org/P5/>, consulté le 4 juin 2004.

<sup>20</sup> <http://www.tei-c.org/Activities/SIG/Manuscript/>