Manuel d'encodage pour les notices épigraphiques du projet TITULUS

Conforme à la TEI P5 et à EpiDoc 8.22

Version 1.1

Élaboré par Clara Renedo Mirambell

Dernière mise à jour : 04/02/2018

Description de l'encodage de la notice

Ces normes ont été conçues spécialement pour le *CIFM*, de sorte qu'elles puissent bien s'adapter à l'encodage de n'importe quelle notice du corpus, et qu'aucune information de la notice ne reste sans avoir été encodée.

Avant de passer à la description individuelle par balises, afin d'avoir une première vue d'ensemble du document entier, nous avons conçu un schéma contenant les balises principales avec des descriptions qui aident à avoir une idée un peu plus précise de la structure générale de la notice encodée :

Structuration de l'encodage

```
<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="schema/tei-epidoc.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>

<!DOCTYPE TEI[
<!ENTITY % epigraphie SYSTEM 'epigraphie.dtd'> %epigraphie;
] >

<TEI xmlns="http://www.tei-c.org/ns/1.0" rend="...">
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title type="inscription">
                        <num n="...">Numéro de l'inscription dans le
                        volume</num>
                        <placeName>Lieu de conservation</placeName>
                        Nom de l'inscription
                  </title>
                <title type="volume">
                        <num>Numéro du volume</num>
                        Corpus des Inscriptions de la France
                        Médiévale
                  </title>
                <author>Nom de l'auteur de l'inscription</author>
                <editor>
```

```
                <persName xml:id="...">
                      <forename>Prénom de l'éditeur</forename>
                      <surname>Nom de l'éditeur</surname>
                </persName>
        </editor>
    <sponsor>
                <orgName xml:id="cescm">CESCM</orgName>
        </sponsor>
    <respStmt>
        <resp when="2016">Encodage en XML-TEI</resp>
        <persName xml:id="...">
                      <forename>Prénom du responsable de
                 l'encodage</forename>
                      <surname>Nom du responsable de
                      l'encodage</surname>
            </persName>
    </respStmt>
</titleStmt>
<publicationStmt>
    <authority>CESCM</authority>
    <address>
        <addrLine>Bât. E13</addrLine>
        <addrLine>24, rue de la chaîne TSA 81118</addrLine>
        <addrLine>86073 Poitiers cedex 9</addrLine>
    </address>
    <idno type="URI">http://cescm.labo.univ-poitiers.fr/
        </idno>
    <date when="2016">2016</date>
    <availability>
        <!-- Type de licence -->
        <licence> </licence>
    </availability>
</publicationStmt>
<sourceDesc>
    <msDesc>
        <msIdentifier>
            <placeName>
                <region type="departement" n="...">
                        Nom du département où se trouve
                        l'inscription
```

```
                    </region>
                <settlement type="city">
                        Nom de la ville
                  </settlement>
                <placeName type="building">
                        Nom de l'édifice
                  </placeName>
                <address>
                    <addrLine>Adresse</addrLine>
                </address>
            </placeName>
<!-- Si l'inscription est dans une institution : -->
            <institution>Nom de l'institution</institution>
            <repository>Lieu de conservation dans
               l'institution</repository>
            <collection>Collection à laquelle appartient
               l'inscription</collection>
            <idno>Numéro d'identification au sein de
               l'institution</idno>
        </msIdentifier>
        <physDesc>
            <objectDesc>
                <supportDesc>
                    <support>
                        <material><term type="...">Type de
                  matériau de support</term>
                            </material>
                        <objectType><term type="mobilier"
                            subtype="...">Objet desupport
                            </term></objectType>
                        <ab>Dimensions du support :
                              <dimensions type="object"
                              unit="cm">
                                <height extent="...">
                                          hauteur
                                    </height>
                                <width extent="...">
                                          largeur
                                    </width>
                                <depth extent="...">
```

```
                                        profondeur
                                        </depth>
                                </dimensions>
                                  </ab>
                        </support>
                        <condition>
                            État de conservation de l'inscription.
                      </condition>
                    </supportDesc>
                    <layoutDesc>
                        <layout>
                            Description de la surface d'écriture.
                               <dimensions type="field" unit="cm">
                                <height extent="...">
                                        Hauteur du champ épigraphique
                                    </height>
                                <width extent="....">
                                      Largeur du champ épigraphique
                               </width>
                            </dimensions>
                        </layout>
                    </layoutDesc>
                </objectDesc>
                <scriptDesc>
                    <scriptNote> Description de l'écriture.
                      <dimensions type="letter" unit="cm">
                            <height extent="...">
                                  Hauteur moyenne des lettres
                               </height>
                            <width extent="....">
                                  Largeur moyenne des lettres
                                </width>
                        </dimensions>
                    </scriptNote>
                </scriptDesc>
                <decoDesc>
                    <decoNote>
                            <term type="iconographie" subtype="...">
                      Description de l'iconographie par mots-clés
                      </term>
```

```
                    </decoNote>
                </decoDesc>
            </physDesc>

            <history>
                <origin>
                    <origPlace>
                    Lieu d'origine de l'inscription
                    </origPlace>
                  <origDate evidence="..." when="...">
                    Datation de l'inscription
                    </origDate>
                </origin>
                <provenance>
                    Situation et propriétaires de l'inscription
                après de son origine mais avant de sa
                situation actuelle.
                    </provenance>
                    <acquisition>
                    Situation actuelle de l'inscription
                    </acquisition>
            </history>
            <additional>
                <adminInfo>
                    <note>Information sur la date et la modalité
                 de consultation de l'inscription.</note>
                </adminInfo>
            </additional>
        </msDesc>
    </sourceDesc>
</fileDesc>
<profileDesc>
    <langUsage>
        <language ident="lat">Langue du texte épigraphique.
          </language>
    </langUsage>
    <textClass>
        <keywords>
            <term type="genre" subtype="...">
                Genre de l'inscription
```

```
                    </term>
                </keywords>
                <!-- Liste des personnages des inscriptions du CIFM -->
            &index-personnages;
                  <keywords>
                 <!-- Liste des mots lemmatisés -->
                  &lexique;
                  </keywords>
            </textClass>
        </profileDesc>
        <encodingDesc>
            <projectDesc>
                <!-- Brève description du projet -->
                <p><rs type="project" xml:id="cifm">
                  Corpus des Inscriptions de la France Médiévale</rs></p>
                <p>Encodage en XML/TEI</p>
            </projectDesc>
            <editorialDecl>
                <p>Déclaration des pratiques éditoriales</p>
            </editorialDecl>
            <charDecl>
                <!-- Déclaration de la liste de glyphes →
                  &allographes;
              </charDecl>
        </encodingDesc>
        <revisionDesc>
            <change when="0001" who="name">
            Description de la revision effectuée à l'encodage
            </change>
        </revisionDesc>
    </teiHeader>

    <facsimile>
        <!-- URI du fichier de l'image -->
        <graphic url="images/CIFM-26_102.jpg">
            <desc>
                <roleName type="author">Nom de l'auteur</roleName>
                <date type="origin">Date de création</date>
                <desc>Description de l'image</desc>
            </desc>
```

```
        </graphic>
    </facsimile>

    <text>
        <body>
            <div type="edition">
                <ab>Encodage du texte épigraphique</ab>
            </div>
            <div type="translation" resp="#..." xml:lang="fr">
                <p>Traduction du texte au français</p>
            </div>
            <div type="commentary">
                <p>Commentaire à l'inscription. </p>
            </div>
            <div type="bibliography">
                <listBibl type="inscription">
                    <!-- Liste d'items bibliographiques  -->
                    <bibl>
                        <author>Auteur</author>
                        <title type="...">Titre de l'ouvrage</title>
                        <date when="1927">Date</date>
                        <biblScope unit="page">Page citée</biblScope>
                       </bibl>
                </listBibl>
            </div>
        </body>
    </text>
</TEI>
```

Les normes d'encodage seront décrites balise par balise en ordre hiérarchique et d'imbrication, sauf pour l'encodage du texte épigraphique, pour lequel l'ordre et la structuration des balises est beaucoup plus complexe et variable, et donc son encodage sera décrit en suivant une structure de caractère thématique.

Le nom de chaque balise va être accompagné du niveau auquel elle se situe par rapport a l'élément racine `<TEI>` : de cette manière, l'élément `<TEI>` va être considéré comme le niveau 1, ses enfants comme le niveau 2, les enfants de ses enfants comme le niveau 3... Cela devrait aider le lecteur à se repérer dans une structure d'une complexité considérable.

D'autre part, il faut indiquer que, si tous les enfants d'un certain élément sont décrits plus haut

et qu'il n'y a pas de caractéristiques spécifiques à remarquer pour l'élément parent, nous garderons le titre avec le niveau et le nom de la balise pour conserver la structure du document, mais il n'y aura pas d'explication particulière propre a l'élément.

Niveau 1 : <TEI> - *Ensemble du document*

Niveau 2 : <teiHeader> - *Informations éditoriales et description de l'inscription*

Niveau 3 : <fileDesc> - *Description bibliographique du document*

Niveau 4 : <titleStmt> - *Titre et auteurs de la notice*

Niveau 5 : <title> - *Titre de la notice*

Le modèle de TITULUS devra contenir, en fait, deux balises `<title>`, l'une pour le titre de l'inscription et l'autre pour le titre du volume. Elles seront distinguées par l'attribut `@type`, ayant comme valeur `'inscription'` pour la première et `'volume'` pour la seconde.

Titre de l'inscription : `<title type='inscription'>`

Cette balise devra contenir les informations suivantes :

- Le **numéro de l'inscription** dans le volume auquel elle appartient, dans la balise `<num>`.
- Un **identifiant unique**, différent pour chaque notice du corpus, à insérer comme valeur de l'attribut `@n` (« numéro »). Il sera composé par le numéro du volume et le numéro de l'inscription dans le volume, en trois chiffres. Ainsi, par exemple, pour l'inscription 3 du volume 24, l'identifiant serait 24003, et pour l'inscription 158 du volume 25, 25158. Cela servira à donner un identifiant unique à chacune des notices du corpus, ce qui est indispensable pour l'indexation.

  Pour les volumes hors-série, il faudra insérer les initiales « HS » avant du numéro propre au volume hors-série. Par exemple, pour l'inscription 12 du volume hors-série II, l'identifiant sera HS2012.
- Le **nom du lieu** où l'inscription est conservée, tel qu'il se trouve dans le titre de l'inscription, dans la balise `<placeName>`.

- Après toutes ces informations, et directement dans la balise `<title>`, **le nom descriptif de l'inscription** en soi, qui en explicite souvent la fonction (tel qu'il apparaît dans le volume papier).

```
<title type="inscription">
      <num n="HSII089">89</num>
      <placeName>Notre-Dame-la-Grande </placeName>
      L'Agneau de Dieu et le tétramorphe
</title>
```

Titre du volume : `<title type='volume'>`

Cette balise devra contenir les informations suivantes :

- Le numéro du volume auquel appartient l'inscription, dans la balise `<num>`.
- Le nom du volume, inséré directement dans le contenu de la balise, et qui va être toujours le même : « Corpus des Inscriptions de la France Médiévale ».

```
<title type="volume">
      <num>HS II</num>
      Corpus des Inscriptions de la France Médiévale
</title>
```

Niveau 5 : <author> - *Auteur de l'inscription*

Cette balise est destinée à contenir le nom de l'auteur de l'inscription (s'il est connu, car le plus souvent il ne l'est pas). Il devra être encodé dans la balise `<persName>`. Les types d'auteur possibles (définis dans le chapitre consacré à l'indexation[1]) devront être encodés dans l'attribut `@role`.

Niveau 5 : <editor> - *Auteur de la notice et de l'édition*

Ici il faudra encoder le nom du responsable de la notice et de l'édition. Il sera encodé dans `<persName>`, avec l'attribut `@xml:id`, ayant comme valeur les initiales de l'éditeur (par exemple, « EIV » pour « Estelle Ingrand-Varenne », ou « RF » pour « Robert Favreau »). Cela peut être utile pour signaler des modifications ou des apports faits par l'éditeur ailleurs dans la notice.

Ensuite, toujours dans la balise `<persName>`, le prénom de l'éditeur sera inséré dans la balise

1 Ils sont : artiste, commanditaire, donateur, fondateur, constructeur, propriétaire.

<forename>, et le nom, dans <surname>.

Dans le cas où il y aurait plusieurs éditeurs, il suffirait de répéter cette tâche, à chaque fois dans une nouvelle balise <persName>.

```
<editor>
      <persName xml:id="EIV">
            <forename>Estelle</forename>
            <surname>Ingrand-Varenne</surname>
      </persName>
</editor>
```

Niveau 5 : <sponsor> - *Organisation responsable de l'édition*

Cette balise devra contenir le nom de l'organisation responsable de l'édition. Pour TITULUS, il s'agira toujours du CESCM. Elle peut être utile si jamais les notices encodées sont utilisées en dehors du projet TITULUS.

Niveau 5 : <respStmt> - *Autres responsables de l'édition*

Cette balise sert à encoder les informations concernant d'autres types de responsabilité relative à l'édition. Dans le cas du *CIFM*, il sera utile de l'utiliser surtout pour décrire et référencer les tâches d'encodage de chaque notice, comme dans l'exemple qui suit :

```
<respStmt>
      <resp  when="2016">Encodage en TEI</resp>
      <persName xml:id="CRM">
            <forename>Clara</forename>
            <surname>Renedo Mirambell</surname>
      </persName>
</respStmt>
```

Elle consiste en deux parties : une première partie encodée dans la balise <resp>, qui sert à définir la responsabilité (en spécifiant la date à travers de l'attribut @when) ; et une deuxième partie avec l'encodage du nom de la personne responsable, qui se fait de la même manière que pour l'auteur (dans <persName>, avec identifiant et contenant les balises <forename> et <surname>).

Si plus tard l'encodage du texte est révisé, si la révision a été faite par le même responsable de l'encodage, dans la même balise `<respStmt>` il faudra ajouter une nouvelle responsabilité avec la balise `<resp>` (et l'attribut `@when` avec la date) :

```
<respStmt>
      <resp when="2016">Encodage en TEI</resp>
      <resp when="2017">Révision de l'encodage en TEI</resp>
      <persNamexml:id="CRM">
            <forename>Clara</forename>
            <surname>Renedo Mirambell</surname>
      </persName>
</respStmt>
```

Si, par contre, il s'agit d'une autre personne, il faudra insérer une nouvelle balise `<respStmt>` en bas de la première et la remplir avec la responsabilité et la date (`<resp>` avec l'attribut `@when`), et le nom du responsable (`<persName>`), avec l'identifiant correspondant (`@xml:id`)[2] :

```
<respStmt>
      <resp  when="2016">Encodage en TEI</resp>
            <persName xml:id="CRM">
                  <forename>Clara</forename>
                  <surname>Renedo Mirambell</surname>
            </persName>
</respStmt>
<respStmt>
      <resp  when="2017">Révision de l'encodage en TEI</resp>
      <persName xml:id="EIV">
            <forename>Estelle</forename>
            <surname>Ingrand-Varenne</surname>
      </persName>
</respStmt>
```

2 Ou `@corresp=#...` si le responsable était déjà apparu plus haut dans l'encodage.

Niveau 4 : <publicationStmt> - *Informations à propos de la publication*

Les informations contenues dans cette balise concernent la publication et ses responsables, et en principe elles devraient être les mêmes pour toutes les notices. C'est pour cela que nous avons pensé à créer une entité contenant tout l'encodage de la balise `<publicationStmt>` qui soit commune à tous les documents. Ce sera une manière d'alléger la tâche d'encodage, et d'éviter de faire des erreurs.

Elle contient les balises suivantes :

- **<authority>** : cette balise est destinée à accueillir le nom de l'autorité responsable de la publication de la notice, qui, pour TITULUS, est le CESCM.

```
<authority>&cescm;</authority>
```

- **<address>** : cette balise doit contenir l'adresse de l'autorité, soit l'adresse du CESCM (organisée par lignes, chacune dans une balise `<addrLine>`).

```
<address>
    <addrLine>Bât. E13</addrLine>
    <addrLine>24, rue de la chaîne TSA 81118</addrLine>
    <addrLine>86073 Poitiers cedex 9</addrLine>
</address>
```

- **<idno>** : cette balise représente un identifiant qui, grâce à l'attribut `@type='URI'`, est utilisé pour indiquer le site web de l'institution (du CESCM) :

```
    <idno type="URI">http://cescm.labo.univ-poitiers.fr/</idno>
```

- **<date>** : cette balise doit contenir la date de l'édition électronique.

```
    <date when="2016">2016</date>
```

- **<availability>** : cette balise contient des informations sur la disponibilité du texte et sa licence de distribution. Dans ce cas-ci, nous avons décidé de le distribuer sous une licence Creative Commons mais qui n'autorise pas à modifier le contenu ou à l'utiliser pour des fins commerciaux.

```
<availability>
 <licence target='https://creativecommons.org/licenses/by-nc-
nd/2.0/fr/'><p>Licence Creative Commons  Attribution - Pas
d'Utilisation Commerciale - Pas de Modification 2.0 France (CC BY-
NC-ND 2.0 FR)</p>
 </licence>
</availability>
```

Niveau 4 : <sourceDesc> - *Description de la source (l'inscription)*

Niveau 5 : <msDesc>

Cette balise contient les informations principales concernant la description bibliographique de l'inscription : les coordonnées sur son lieu de conservation, la description physique, l'histoire et la date et le moyen de consultation.

Niveau 6 : <msIdentifier> - *Données sur la conservation de l'inscription*

Cette balise contient des informations concernant le lieu et l'autorité responsable de la conservation de l'inscription dans l'actualité.

Niveau 7 : <placeName> - *Lieu géographique*

Il s'agit du lieu où l'inscription se trouve actuellement, même s'il est différent du lieu d'origine. Comme le *CIFM* ne concerne que le territoire français, cela nous permet d'homogénéiser énormément les informations concernant la situation de l'inscription. Ainsi, les catégories que la balise `<placeName>` devrait contenir sont les suivantes :

- Département (dans la balise `<region type='department' n='...'>`, où l'attribut `@n` devra contenir comme valeur le numéro de département) ;
- Commune (dans la balise `<settlement @type='city'>`) ;
- Si l'inscription se trouve dans un édifice, il faudra aussi indiquer son nom dans la balise `<placeName type='building>` ;
- Finalement, l'adresse (dans la balise `<address>`, et organisée par lignes grâce à `<addrLine>`), avec la localisation précise sur le site, contenant le nom et le numéro de la rue, le code postal et la ville.

En voici un exemple :

```
<placeName>
      <region type="departement" n="86">Vienne</region>
      <settlement type="city">Poitiers</settlement>
      <placeName type="building">Cathédrale</placeName>
      <address>
            <addrLine>1 rue Sainte-Croix</addrLine>
            <addrLine>86000 Poitiers</addrLine>
      </address>
```

```
    </placeName>
```

Niveau 7 : <institution>, <repository>, <collection>, <idno> - ***Autres informations concernant l'appartenance à une institution***

Si l'inscription appartient à une institution, il faudra rentrer le nom de l'institution dans la balise `<institution>`.

Il faut tenir compte que l'institution n'est pas forcément la même que le lieu (l'édifice) où l'inscription est conservée, bien que ce soit souvent le cas. L'institution est l'organisation à laquelle l'inscription appartient du point de vue administratif, mais elle peut être conservée ailleurs (par exemple, dans les magasins du musée). L'information à propos du lieu de conservation va être encodée dans la balise `<repository>`. Si l'inscription appartient à une collection, cette information devra rentrer dans la balise `<collection>`. Pour le numéro d'inventaire, la balise à utiliser est `<idno>`.

En voici un exemple :

```
    <institution>Musée du Berry</institution>
    <repository>réserves Saint-Aoustrillet</repository>
    <collection/>
    <idno>1903.5.1</idno>
```

Niveau 6 : <physDesc> - ***Description matérielle du support, du texte et de l'écriture***

Description du support et de la surface d'écriture (`<objectDesc>`), ainsi que de l'écriture elle-même (`<scriptDesc>`).

Niveau 7 : <objectDesc> - ***Description complète du support et de la surface d'écriture***

Description du support de l'inscription (`<supportDesc>`) et de la surface d'écriture (`<layoutDesc>`).

Niveau 8 : <supportDesc> - ***Description du support et son état de conservation***

Description du support (dans la balise `<support>`) ainsi que de son état de conservation (dans la balise `<condition>`).

Niveau 9 : <support> - ***Description du support***

L'élément `<support>` contient les informations principales qui servent à décrire le support, organisées dans les balises `<objectType>`, `<material>` et `<dimensions>`. Dans le cas (très rare)

où l'inscription se trouverait dans plusieurs supports, il faudrait prévoir une balise `<support>` pour chaque support.

- **`<ObjectType>`** : Type d'objet ou monument de support. Afin d'en optimiser l'indexation, il sera utile d'insérer dans `<objectType>` une balise `<term>` avec attribut `@type`, qui définit la catégorie dans laquelle rentre le matériau. Comme nous l'avons vu dans le chapitre consacré à l'indexation, les inscriptions médiévales du *CIFM* sont classées selon trois types d'objets : éléments d'architecture (tout simplement `'architecture'`), mobilier (comme `'mobilier'`) et objet (`'objet'`). Ces catégories peuvent avoir des sous-catégories, qui seront à encoder dans un autre attribut `@subtype`. Pour les sous-catégories composées de plusieurs mots, il faudra substituer les espaces (qui ne sont pas autorisés) par des tirets.

Valeurs possibles des attributs `@type` et `@subtype` de la balise `<support>`

Catégorie	Valeur de @type correspondant	Sous-catégorie	Valeur de @subtype correspondant
Architecture	`'architecture'`	Tympan	`'tympan'`
		Clé du linteau	`'clé-du-linteau'`
		Chapiteau	`'chapiteau'`
		...	
Mobilier	`'mobilier'`	Retable	`'retable'`
		Plaque funéraire	`'plaque-funéraire'`
		Plate-tombe	`'plate-tombe'`
		...	
Objet	`'objet'`	Anneau	`'anneau'`
		Croix-reliquaire	`'croix-reliquaire'`
		Dé à jouer	`'dé-à-jouer'`
		...	

Voici un exemple d'encodage d'<objectType> :

```
<objectType>
      <term type="architecture" subtype="clé-du-linteau">
      clé du linteau
      </term>
</objectType>
```

- **<material>** : Matériau. Ici aussi, pour en faciliter l'indexation, on introduira dans la balise <material> une balise <term> avec attribut @type, qui définit la catégorie dans laquelle rentre le matériau. Dans l'attribut @type, les valeurs possibles vont être les catégories déjà décrites dans le chapitre à propos de l'indexation (pierre, métal, bois, céramique, émail, peintures murales, textile et verre). Si le matériau spécifié dans la notice est plus concret (par exemple, il est indiqué « argent », et non pas simplement « métal »), on pourra rajouter cette précision avec l'attribut @subtype. Pour les catégories composées de plusieurs mots, telles que les peintures murales, il suffira d'ajouter un tiret entre chaque mot : 'peintures-murales'.

Valeurs possibles des attributs @type et @subtype de la balise <material>

Catégorie	Valeur de @type correspondant	Sous-catégorie	Valeur de @subtype correspondant
Pierre	'pierre'	Marbre	'marbre'
		Ardoise	'ardoise'
		Calcaire	'calcaire'
		...	
Métal	'métal'	Argent	'argent'
		Plomb	'plomb'
		...	
Bois	'bois'	- - -[3]	

3 Le fait qu'il n'y ait pas de sous-catégorie définie pour les autres types de matériau tels que le bois ne veut pas dire qu'il ne soit pas possible de les inclure. C'est simplement que dans les volumes papier et dans les notices

Céramique	'céramique'	---	
Peintures murales	'peintures-murales'	---	
Textile	'textile'	---	
Verre	'verre'	---	

- **<dimensions>** : Dimensions de l'objet. Cette balise devra contenir par défaut deux attributs avec des valeurs figés : `type="object"` et `unit="cm"`. Le premier sert à identifier que les dimensions sont bien celles de l'objet épigraphique (car, comme nous le verrons plus avant, il sera possible d'indiquer aussi les dimensions de la surface d'écriture ou même des lettres), et le deuxième indique que les dimensions seront exprimées en centimètres (dans les cas où une autre mesure est utilisée dans la notice, il faudra faire la conversion, afin de s'assurer de l'homogénéité de cette information dans toutes les notices du corpus). Dedans, se trouvent trois balises pour les trois dimensions : `<depth>`, `<width>` et `<extent>`. Comme parfois la valeur numérique peut être accompagnée de texte qui donne des précisions, nous avons décidé d'inclure l'attribut `@extent`, qui aura comme valeur uniquement la valeur numérique de la balise correspondante. Si la valeur d'une des balises nous est inconnue (comme souvent pour `<depth>`), il s'agira de remplir l'attribut `@extent` avec la valeur `'unknown'`. D'autre part, il ne faudra pas répéter à chaque balise que les données sont en centimètres, car cela est déjà spécifié dans la balise `<dimensions>`.
  Si les dimensions de l'objet nous sont complètement inconnues, il suffira d'ajouter directement l'attribut `@extent` et la valeur `'unknown'` dans la même balise `<dimensions>` :

```
<dimensions type="object" unit="cm" extent="unknown"/>
```

Voici un exemple de description complète de la balise `<support>` :

```
<support>
      <objectType>Plaque</objectType> de
      <material>plomb</material> de
      <dimensions type="object" unit="cm">
```

observées cette information n'est jamais précisée.

```
                <depth extent="unknown"/>
                <width extent="16">16</width>
                <height extent="6">6</height>
            </dimensions>
        </support>
```

Niveau 9 : <condition> - *État de conservation*

On ne vise pas l'homogénéisation, car il s'agit d'une information qui est peut-être assez subjective, et, par ailleurs, il est très difficile de comparer entre l'état de conservation de différents objets, techniques et supports.

D'autre part, il s'agit d'une information qui a un intérêt relatif pour l'indexation, car il n'est pas une caractéristique de l'inscription en soi, mais une conséquence de son évolution.

Il s'agira donc juste de remplir cette balise avec les informations présentes dans la notice concernant l'état de conservation de l'inscription.

Niveau 8 : <layoutDesc> - *Description de la mise en page et de la technique d'exécution*

La description de la disposition de l'écriture sur le support et la technique d'exécution est à insérer dans la balise `<layout>`, contenue dans `<layoutDesc>`. D'après ce que nous avons observé dans notre corpus de notices, il s'agit d'une description assez libre qui peut varier énormément d'une notice à l'autre (en fonction des caractéristiques de l'inscription) : il sera donc inutile d'essayer de la systématiser.

Si l'inscription est divisée en plusieurs parties physiques qui ne sont pas homogènes, et que la description est faite en plusieurs parties bien distinctes, il sera utile de diviser le contenu en plusieurs balises `<layout>`.

Niveau 9 : <scriptDesc> – *Description paléographique*

Comme pour la description de la mise en page, ici il s'agit aussi d'informations assez variables en fonction des particularités de l'inscription.

D'autre part, comme nous le verrons plus avant, le contenu de la description paléographique, même s'il n'est pas directement systématisé dans `<scriptDesc>`, va être une source fondamentale

d'information pour l'encodage des informations paléographiques dans le texte épigraphique lui-même.

Niveau 9 : <decoDesc> – *Description de l'iconographie et la décoration*

Il s'agit ici d'un cas très particulier, car ce type d'information n'apparaît pas dans la notice de manière systématique (elle est mentionnée parfois dans le commentaire général). La raison en est qu'il ne s'agit pas d'un corpus à visée iconographique, et comme le lecteur dispose déjà d'une image photographique de l'inscription pour apprécier lui-même l'iconographie qui entoure l'inscription avec plus de détail, il n'y a pas besoin d'en faire une description exhaustive. Mais, comme nous l'avons vu dans le chapitre consacré à l'indexation, des références à l'iconographie accompagnant les inscriptions apparaissent dans les index thématiques.

Afin de maintenir et améliorer ce travail d'indexation qui avait déjà été fait pour l'édition papier, nous avons décidé de baliser ces informations d'une manière assez simple et basique. Comme nous l'avons déjà expliqué dans l'indexation, le travail de description sera simplifié dans une simple liste de mots clés de personnages, éléments ou motifs. Ainsi, dans la balise `<decoNote>`, il faudra insérer cette information dans des balises `<term>` (« terme »), avec l'attribut `@type='iconographie'`, et l'attribut `@subtype` dont la valeur sera le terme iconographique désiré (toujours en substituant les espaces par des tirets). Dans le contenu de la balise, nous pourrons inclure une brève description du terme iconographique :

```
<decoDesc>
   <decoNote>
         <term type='iconographie' subtype='terme-iconographique'>
         Description du terme iconographique
         </term>
    </decoNote>
  </decoDesc>
```

Mais, afin de nous assurer que les termes utilisés coïncident lorsque les termes décrits sont les mêmes, il sera utile de les encoder comme des entités dans un document externe. Ainsi, au fur et à mesure on assistera à la création d'un répertoire d'éléments iconographiques qui devra faciliter l'encodage et l'indexation de ces termes.

Niveau 6 : <history> - *Histoire de l'inscription*

La description de l'histoire de l'inscription est divisée en trois périodes :

- Informations concernant l'origine de l'inscription, à encoder dans la balise `<origin>` et dans ses sous-éléments :
  - `<origPlace>`, pour le lieu d'origine.
  - `<origDate>`, pour la date d'origine. Les informations concernant la date sont à insérer dans la balise, mais il sera intéressant également de les systématiser d'une manière compréhensible pour la machine, qui en facilite l'indexation, à travers des attributs :
    - Attribut `@evidence` : à travers des valeurs déjà prédéfinies, il sert à indiquer la manière dont la datation a été établie. Cette information est toujours donnée dans la notice.
    - Attributs concernant la date : si l'année précise est connue, il faudra utiliser l'attribut `@when`, avec la valeur numérique de l'année. En revanche, si l'information n'est pas assez précise, et qu'on n'a pas affaire à une date concrète, mais à un intervalle (ce qui arrive très souvent), il conviendra d'utiliser balises `@notBefore` et `@notAfter`. Par exemple, si la datation est fixée à la « deuxième moitié du XIe siècle », l'encodage sera : `@notBefore='1051' @notAfter='1100'`.

      Important : la date doit impérativement être composée de quatre chiffres. Si elle est inférieure à l'an 1000, il faudra donc insérer un zéro à gauche. Par exemple, pour l'année 950 : `@when='0950'`.
- Informations concernant la période comprise entre l'origine et l'état actuel, dans la balise `<provenance>`. Si cette période est divisée en plusieurs étapes, il faudra utiliser une balise `<provenance>` pour chacune d'elles. Il sera utile de se servir des attributs `@when`, `@notBefore` ou `@notAfter` pour encoder la date.
- Informations concernant l'état actuel de l'inscription, et surtout sa localisation. Nous avons décidé de nous servir ici de la balise `<acquisition>` (« acquisition »). Cette balise était pensée originellement pour décrire les livres, qui ont toujours un propriétaire, et elle servait à décrire la dernière acquisition du livre en question (c'est-à-dire, l'acquisition par laquelle le livre était arrivé à son propriétaire actuel). Pour les inscriptions, même si cela peut être aussi le cas, ce sont assez nombreuses les inscriptions qui ont tout simplement changé de localisation dans le même édifice, par exemple, et donc le dernier chapitre de leur histoire n'est pas une

acquisition à proprement parler. Cependant, l'élément précédent, <provenance>, ne conviendrait pas non plus pour l'encodage de ce type de cas, d'après sa définition il paraît assez clair qu'il sert à décrire des étapes intermédiaires de l'histoire du texte, après son origine mais avant son état actuel. Après y avoir longuement réfléchi, nous avons décidé de garder cette balise même dans ces derniers cas, car, même si l'identification avec le sens de la balise n'est pas complet, c'est celui qui s'y rapproche le plus.

Voici un exemple complet d'encodage de l'élément <history>, appartenant à l'inscription 157 du volume 26 du *CIFM* :

```
<history>
  <origin>
      <origPlace>Localisation originelle inconnue.</origPlace>
      <origDate evidence="lettering" notBefore="1126"
      notAfter="1150">deuxième quart du XII&e-super; siècle
      [datation stylistique en accord avec l'analyse
      paléographique].</origDate>
  </origin>
  <provenance notAfter="1900">Avant 1900 elle se trouvait insérée
  dans le mur oriental du cloître, c'est-à-dire le mur ouest du bras
  sud du transept.</provenance>
  <acquisition notBefore="1900">Localisation actuelle : intérieur,
  transept, bras sud, mur est, à 1,53 m du sol.</acquisition>
</history>
```

Niveau 6 : <additional> - *Information bibliographique additionnelle*

Niveau 7 : <adminInfo> - *Information sur les modalités d'accès à l'inscription*

Nous avons encodé ici les informations concernant la date et la modalité de consultation de l'inscription, à partir de laquelle l'éditeur s'est fondé pour rédiger la notice. Il s'agit d'une information qui est toujours présente dans les notices du *CIFM*.

Elle devra être encodée dans la balise <note>. En voici l'encodage pour la notice 5 du vol. 25 :

```
<additional>
   <adminInfo>
        <note>D'après l'original vu au musée en 1994. </note>
```

```
    </adminInfo>
  </additional>
```

Niveau 3 : <ProfileDesc> - *Description non-bibliographique de l'inscription*

Cette balise regroupe toute sorte d'informations par rapport à l'inscription qui ne sont pas de type bibliographique, tels que la langue utilisée ou les mots clés.

Niveau 4 : <langUsage> - *Usage des langues*

Cette balise est destinée à encoder la langue ou les langues dans lesquelles est écrite l'inscription. Il s'agira d'encoder chaque langue dans une balise `<language>`, avec l'attribut `@ident` qui devra contenir le code de la langue attribué par IANA[4]. Ce code est `'la'` pour le latin et `'fr'` pour le français :

```
    <language ident="la">Latin</language>
    <language ident="fr">Français</language>
```

Niveau 4 : <textClass> - *Classification de termes décrivant la nature ou sujet du texte*

Cette balise est réservée à la classification du texte par des termes issus d'un système standardisé.

Dans TITULUS, elle va être utilisée pour décrire le genre du texte épigraphique, et accueillir la liste des lemmes, l'index des lieux et l'index des personnages.

Description du genre du texte

Le genre ou fonction de l'inscription est une information qui est toujours mentionné dans la notice, mais non de manière systématique. Comme nous avons vu dans l'indexation, une information très similaire apparaît de manière beaucoup plus succincte et systématisée dans l'index thématique des volumes. Il s'agira ici de faire une combinaison des deux, afin de pouvoir indexer l'information sans perdre sa richesse.

Nous avons opté pour maintenir la description du genre présente dans la notice papier exactement telle qu'elle est, et faire le travail de systématisation dans le balisage. En premier lieu, le terme va être encodé dans la balise `<term type='genre'>`. Deuxièmement, il faudra compléter cette balise avec un autre attribut, `@subtype`, destiné à contenir genre principal de l'inscription. Si la

4 IANA (Internet Assigned Numbers Authority) est une organisation privée américaine sans but lucratif qui est en charge de l'attribution des adresses IP et de tous les codes en relation avec le protocole d'internet.

description contient des informations qui peuvent être considérées comme des sous-genres (en nous fondant sur la systématisation des termes déjà faite dans le chapitre réservé à l'indexation), nous pourrons baliser ces informations avec de nouvelles balises `<term>`, tout en définissant la catégorie dans une balise `@type` (il ne faudra pas répéter la valeur `@type='genre'`, car cette nouvelle balise `<term>` sera un enfant de la première balise qui contenait déjà cet attribut).

Dans cette table, conçue à partir des résultats de la réflexion sur l'indexation menés dans le chapitre antérieur, nous pourrons comprendre beaucoup mieux le mécanisme de balisage des genres et des sous-genres :

GENRE PRINCIPAL		SOUS-GENRES		
Nom	**Valeur dans @subtype**	**Nom**	**Balise enfant <term>**	
			Valeur dans @type	**Valeur dans @subtype**
Campanaire	`campanaire`			
Commentaire iconographique	`commentaire-image`			
Concession de droit	`concession-droit`			
Datation	`datation`			
Donation	`donation`			
En lien avec la liturgie	`liturgique`	Consécration (d'autel)	`consécration`	`autel`
		Dédicace (d'église)	`dédicace`	`église`
Fondation	`fondation`			
Funéraire	`funéraire`	Déplacement de sépulture	`déplacement-sépulture`	
		Fondation d'anniversaire	`fondation-anniversaire`	
		Formule d'absolution	`formule-absolution`	
		Laudative	`laudative`	
		Littéraire	`littéraire`	
		Obituaire	`obituaire`	

		Tumulaire	tumulaire	
		Prosopographie	prosopographie	abbé / prêtre / laïc / roi / chevalier ...
		...		
		Indéterminée	---	
Identification	identification	Lieu	lieu	
		Mois	mois	
		Objet	objet	
		Personnage	personnage	
		Scène	scène	
		...		
Prophylactique	prophylactique			
Inventaire de reliques	inventaire-reliques			
Invocation	invocation			
Limite de mur	limite-mur			
Mention de construction	mention-construction			
Prière	prière			
Recommandation morale	recommandation-morale			
Signature	signature			
Vœu pieux	vœu-pieux			
Indéterminée	indéterminée			

Voici quelques exemples d'encodage en suivant cette table :

```
    <term type="genre" subtype="funéraire">Inscription funéraire pour un
<term type="prosopographie" subtype="prêtre">prêtre</term></term>

    <term type="genre" subtype="commentaire-image">Commentaire des
scènes de la vie et la mort de la Vierge</term>
```

```
    <term type="genre" subtype='identification'>Identification de <term
type="personnages">l'Agneau et des évangélistes</term></term>
```

Cette table contient une première liste des genres et sous-genres des inscriptions, élaborée à partir des informations extraites des index des volumes 24-26 et HS II, mais rien n'empêche que d'autres genres ou sous-genres puissent apparaître, ou, ce qui est plus probable, qu'ils apparaissent en combinaison. Voici un exemple d'encodage d'inscription funéraire qui combine deux sous-genres (appartenant à la notice 5 du volume 25 du *CIFM*) :

```
     <term type="genre" subtype="funéraire">
          Inscription funéraire
          <term type="tumulaire">à caractère tumulaire</term>
          pour un
          <term type="prosopographie" subtype="prêtre">prêtre</term>
    </term>
```

Lemmatisation

Comme nous le verrons plus avant dans l'encodage du texte épigraphique, les mots du texte vont être tous lemmatisés. La liste des lemmes va être encodée ici, à travers une entité (qui permettra que cette liste soit la même pour toutes les inscriptions).

Index des lieux et personnages

Nous verrons également dans l'encodage du texte épigraphique la manière d'encoder les références aux lieux et personnages qui apparaissent nommés dans l'inscription. Cet encodage va se faire en pointant vers une liste de lieux et personnages insérée ici à travers une entité (afin qu'elle soit partagée par l'ensemble du corpus).

Niveau 3 : <encodingDesc> - *Informations concernant l'encodage de la notice*

Niveau 4 : <projectDesc> - *Description des objectifs de l'encodage*

Cette balise est destinée à contenir une description à propos des objectifs de l'encodage de la

notice. Comme les objectifs sont communs au projet, et donc ils sont les mêmes pour toutes les notices du *CIFM*, il sera très pertinent d'encoder cette information dans une entité. Ainsi, si jamais on décide de réviser cette description, il suffira de faire les changements pertinents une seule fois.

Niveau 4 : <editorialDecl> - *Description des principes éditoriaux de l'encodage*

Cette balise est pensée pour accueillir les principes éditoriaux de l'encodage de la notice. Comme pour la balise antérieure, les principes étant communs à toutes les notices, il sera très utile d'utiliser une entité ayant comme contenu l'encodage de cette balise.

Niveau 4 : <charDecl> - *Déclaration des caractères non-standardisés*

Cette balise sert à définir les caractères non-standardisés qui, ayant une forme qui n'est pas représentée par l'alphabet latin standard, ont dû être encodés pour pouvoir être représentés dans le texte. Nous avons traité avec beaucoup plus de détails l'encodage de cette information dans la description de l'encodage des caractères dans la transcription du texte épigraphique.

Niveau 2 : <facsimile>

Cette balise est destinée à la mise en rapport du texte avec sa représentation graphique. Ici, nous l'utiliserons pour insérer les liens vers les fichiers des images et donner les informations essentielles à propos de chaque image.

Pour chaque image il faudra rajouter une balise `<graphic>`. Le chemin vers le fichier de l'image sera défini dans l'attribut `@url`. Ensuite, il faudra rajouter une balise `<desc>` (« description »), contenant le nom de l'auteur dans une balise `<roleName type='author'>` et la date de création dans une balise `<date type='origin'>`. Si ces informations ne sont pas disponibles, on pourra se passer de ces balises.

Au cas où la notice contient plusieurs images, il est recommandé de rajouter une description succincte de chaque image dans une balise `<desc>` contenue par la balise `<desc>` que nous venons de mentionner.

En voici un exemple :

```
<facsimile>
    <graphic url="images/49.JPG">
        <desc>
```

```
                <roleName type="author"/>
                <date type="origin"/>
                <desc>Détail de la partie inférieure du tympan</des>
            </desc>
        </graphic>
    </facsimile>
```

Niveau 2 : <text>

Niveau 3 : <body> - *Le texte épigraphique à proprement parler*

La particularité d'Epidoc par rapport au schéma général de la TEI est que le `<body>` (destiné à contenir le corps du texte), est divisé en quatre parties par défaut :

- Une première partie réservée à la transcription du texte épigraphique, encodée dans `<div type='edition'>`.
- Une deuxième balise `<div type='translation'>` pour la traduction de l'inscription.
- Une troisième partie réservée à accueillir un commentaire à propos de l'inscription, dans `<div type='commentary'>`.
- Et, enfin, une dernière partie destinée à la bibliographie (`<div type='bibliography'`).

Toutes ces quatre sous-divisions se trouvent aussi dans la notice de l'édition papier du *CIFM*, et donc pour TITULUS nous allons conserver cette structure sans y toucher.

Niveau 4 : <div @type='edition'> - *Encodage du texte épigraphique*

La particularité de l'encodage du texte dans un langage comme la TEI est que la transcription et l'édition normalisée de l'édition papier peuvent s'unir dans un unique texte, tout en fusionnant tous les niveaux de représentation concernés.

Il faudra donc considérer la transcription et l'édition normalisée comme une seule chose, et tenir compte encore une fois qu'on n'encode pas pour rendu, mais pour décrire le texte (même si tout n'est pas visible dans l'affichage, il y aura des informations qui peuvent être intéressantes à encoder et qui peuvent servir à d'autres finalités).

D'autre part, il faudra aussi tenir compte que l'édition reste un choix, un point de vue ; et que

ce point de vue est, à la base, celui défini par l'édition papier du *CIFM*.

L'encodage du texte est, probablement, la partie de TITULUS la plus différente par rapport à EpiDoc, car c'est là où se voient vraiment les différences entre les inscriptions médiévales et les inscriptions classiques. Ainsi, par exemple, EpiDoc a beaucoup développé l'encodage des dégâts et des parties endommagées. En revanche, cela n'a jamais été traité si profondément pour l'épigraphie médiévale dans le *CIFM*, et c'est pour cela qu'il faudra simplifier considérablement cette partie.

D'autre part, des caractéristiques graphiques telles que les jeux de lettres et les glyphes sont presque inexistants ou très simplifiés dans EpiDoc, tandis qu'elles sont très présentes dans les inscriptions médiévales, et par conséquent elles ont suscité l'intérêt des épigraphistes de cette période depuis longtemps.

Par ailleurs, les différences par rapport à EpiDoc ne concernent pas que les particularités de l'épigraphie médiévale, mais elles touchent aussi aux besoins spécifiques de notre projet (comme nous le verrons, par exemple, pour l'indexation ou la forme des mots).

Le but sera toujours de maintenir le même esprit des notices de l'édition papier, tout en enrichissant certaines parties, afin de profiter au maximum des possibilités du format numérique.

Heureusement pour nous, ce travail d'adaptation de l'encodage des textes épigraphiques médiévaux avec EpiDoc ne part pas de zéro, mais il s'appuie sur les solutions adoptées et les réflexions faites pour l'encodage du corpus d'inscriptions épigraphiques médiévales du projet ORIFLAMMS, dirigé par Dominique Stutzmann (IRHT).

Comme nous l'avons déjà vu, ce corpus était constitué également par une sélection de notices du *CIFM*. Au tout début, nous nous sommes demandés s'il ne serait pas pertinent d'adopter exactement les mêmes principes d'encodage et réutiliser les textes qui avaient été déjà encodés.

Mais nous nous sommes très vite rendus compte que l'encodage d'ORIFLAMMS ne convenait pas très bien à TITULUS, car les objectifs des deux projets ne sont pas exactement les mêmes, et ce dernier est beaucoup plus exhaustif. Ce ne sont pas que les formes de l'écriture médiévale qui nous intéressent, mais aussi sa structuration grammaticale et syntaxique, le contenu du texte...

Le modèle d'encodage de TITULUS cherche à poser les bases pour une nouvelle manière de concevoir l'édition des textes épigraphiques médiévaux. Il s'agit d'une évolution qui n'est pas seulement technique, car les améliorations technologiques devront être toujours le résultat des progrès dans l'étude de l'épigraphie médiévale. La vraie révolution ne sont pas uniquement les moyens techniques, mais aussi l'enrichissement et la mise à jour de la discipline.

Compte tenu de la brièveté et l'hétérogénéité des textes épigraphiques, nous avons considéré qu'il était faisable et très intéressant d'opter pour un encodage très riche dans l'édition.

Transcription et interventions éditoriales

Surface de texte

Il s'agit de délimiter l'unité physique du texte telle qu'elle a été conçue. Par exemple, dans la figure 1, chacun des noms des personnages devra être dans une unité de texte différente. Si le texte a été divisé a posteriori, comme dans la figure 2 (où, suite à un accident, la pierre qui porte l'inscription a été coupée en deux), cela n'aura pas d'incidence sur l'encodage de la surface de texte, qui sera considérée comme une même unité.

Figure 1 - *CIFM*, 25, 65, détail. *Cliché CESCM/CIFM*

Figure 2 - *CIFM*, HS II, 2. *Cliché CESCM/CIFM*

Pour le choix de la balise, nous avons décidé d'utiliser toujours l'élément `<ab>`, qui est destiné à contenir « une unité de texte quelconque, [...] analogue à un paragraphe mais sans sa portée sémantique »[5].

Nous avons écarté très vite la balise `<p>` (qui sert à indiquer un paragraphe et qui avait été utilisée au début dans ORIFLAMMS), car on n'a pas toujours affaire à un texte en prose, et, en tout cas, ce n'est pas l'unité sémantique qu'on est en train de décrire ici, mais l'unité physique. Il sera préférable donc de réserver l'élément `<p>` à la fonction qui lui est propre.

Lorsque le texte se divise en plusieurs zones, la transcription peut être précédée d'une brève description de la localisation de chaque zone. Dans ce cas-là, la description devra être encodée dans la balise `<desc>`. Par exemple, pour la figure 1, l'encodage pourrait être le suivant :

5 Définition donnée dans les *TEI Guidelines* (http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-ab.html).

```
<desc>Sur le personnage à gauche de l'arbre :</desc>
<ab>DEVS</ab>
<desc>Sur le personnage à droite de l'arbre :</desc>
<ab>CREATOR</ab>
```

DÉGÂTS ET PARTIES ENDOMMAGÉES

Par rapport aux possibilités d'EpiDoc, la description de cette partie va être beaucoup plus allégée. La raison principale en est que l'intérêt des épigraphistes médiévistes français, comme nous avons vu dans l'édition des volumes du *CIFM*, est tout simplement de signaler les parties perdues en identifiant les caractères concernés, et essayer de les restituer si possible. Dans le commentaire des notices, il est spécifié parfois la raison de la perte, mais normalement il n'y a aucune autre information de ce type.

Nous avons décidé d'ajouter l'encodage des caractères qui n'ont pas été complètement perdus mais qui sont fort endommagés (et dont le déchiffrement se fait en partie par hypothèse).

Texte perdu sans restitution possible

Pour signaler une omission dans une transcription parce que le matériau est illisible, nous avons décidé d'utiliser la balise `<gap>`, qui sert à indiquer « une omission dans une transcription ». Pour TITULUS nous avons décidé de ne l'utiliser que si le texte perdu est impossible à restituer.

Les précisions à faire sont, d'un côté, la quantité de texte perdu, qu'on devra préciser à travers les attributs `@extent` et `@unit` :

- `@unit` sert à définir l'unité de mesure de la quantité de texte perdu. Par défaut, dans le *CIFM* l'unité va être le caractère : `@unit='character'`.
- `@extent` sert à définir la quantité de caractères perdus, exprimés à travers d'une valeur numérique. S'il n'y a qu'un seul caractère perdu, `@extent='1'`. Si la quantité de caractères perdus n'est pas connue, la valeur devra être '`unknown`' (inconnue).

Figure 3 - *CIFM*, 26 97, « DEFERT AD [...] ». *Cliché CESCM/CIFM*

Ainsi, par exemple, le texte perdu de la figure 3 doit être encodé de la manière qui suit :

```
DEFERT AD G<gap unit='character' extent='unknown'/>
```

De l'autre côté, il sera intéressant de préciser aussi les raisons de la perte du texte. Les deux options par défaut (qui ont été réduites par rapport à celles proposées par EpiDoc, tout en supprimant celles qui n'étaient pas adéquates pour cet usage[6]) sont `'illegible'` (au cas où le texte serait encore là mais impossible à interpréter) et `'lost'` (au cas où le texte serait complètement perdu, comme dans la figure 3).

Si le texte perdu a une partie qui peut être restituée, voir « texte perdu mais reconstruit par hypothèse ».

Texte perdu mais reconstruit par hypothèse

Si le texte perdu peut être reconstitué, alors il conviendra d'utiliser la balise `<supplied>`, qui « permet d'encoder du texte restitué par l'auteur de la transcription ou par l'éditeur pour une raison quelconque, le plus souvent parce que le texte du document original ne peut être lu, par suite de dommages matériels ou de lacunes ». Elle est à utiliser de la même manière que `<gap>` pour le texte perdu sans restitution possible, mais ici il s'agira de rétablir les lettres manquantes dans le contenu de la balise. Les attributs seront aussi les mêmes, et ils seront à utiliser exactement de la même manière.

D'autre part, si le même bloc de texte à reconstituer appartient à deux ou plusieurs mots successifs, cela pose un problème d'encodage, vu que, comme nous le détaillerons plus avant, dans TITULUS chaque mot sera encodé individuellement dans une balise `<w>`.

Ainsi, par exemple, dans l'inscription 104 du volume 26, dans la tranche gauche du montant de la croix, le texte manquant est « de » et la première partie du mot « sepulcro ». Théoriquement l'encodage devrait être ce qui suit :

```
<w><supplied>de</w><w>sepul</supplied>cro</w>
```

Cependant, cet encodage ne respecte pas la règle principale du langage XML (l'imbrication des éléments à la manière des poupées russes, dans laquelle il n'est pas possible que les éléments se chevauchent), et donc le document ne serait pas valide. Cela va être un problème assez fréquent dans notre encodage, car, comme il comprend différents niveaux de compréhension du texte tous ensemble, cela va souvent générer des incompatibilités de ce type.

6 Comme « `ellipsis` », qui est réservée au texte omis par l'éditeur. Comme ici il s'agit de transcrire exactement ce qu'il y a exactement dans l'inscription, cette valeur n'avait pas raison d'être.

La solution la plus simple passe par hiérarchiser les balises, et découper en plusieurs parties celles qui se trouvent au niveau le plus bas. Ici, nous avons décidé que c'était le mot (et donc la balise `<w>`) qui avait la priorité, car il s'agit, en quelque sorte, de l'unité minimale de structuration du texte.

Nous subordonnerons l'encodage des parties manquantes aux unités des mots : il s'agira de découper une même balise `<supplied>` en plusieurs sous-parties, mais en faisant le lien entre elles grâce aux attributs `@xml:id` (pour la première balise) et `@corresp` (pour les balises suivantes), toutes avec la même valeur :

```
<w><supplied xml:id='de-sepul'>de</supplied></w>
<w><supplied corresp='#de-sepul'>sepul</supplied>cro</w>
```

Texte perdu et reconstituable seulement en partie

Si le texte n'est reconstituable qu'en partie, la solution sera d'utiliser une combinaison des deux options antérieures : il s'agira d'encoder la partie complètement perdue avec la balise `<gap>`, et la partie perdue mais reconstituable avec la balise `<supplied>`. Les informations sur la quantité de caractères perdus seront à ajouter dans chacune des balises, avec la quantité correspondante dans chaque cas (par exemple, si nous avons une partie de texte perdu qui se compose par une quantité de caractères indéfinie qui n'est pas reconstituable, et trois caractères qui sont reconstituables, la première partie devra être encodée telle qu'elle suit, `<gap unit="character" extent="unknown"/>`, et pour la deuxième ce sera `<supplied unit="character" extent="3">`*(caractères reconstitués)*`</supplied>`).

Cependant, afin de pouvoir reconnaître qu'il s'agit bien du même ensemble de texte perdu, il conviendra d'établir une correspondance en leur attribuant un même identifiant (à travers des attributs `@xml:id` et `@supplied`).

Texte sérieusement endommagé, mais conservé et lisible

Cela n'avait jamais été signalé dans la transcription de l'édition papier du *CIFM*. La raison principale en est que, comme le rendu du texte endommagé dans l'édition doit être une convention (car on ne peut pas reproduire le dommage), les éditeurs considéraient que cela rendait la transcription visuellement trop codifiée, avec trop de symboles arbitraires, ce qui en rendait plus difficile la lecture.

Ici, vu que le rendu n'est pas la finalité immédiate de l'encodage, et qu'on encode tout d'abord

pour décrire, nous avons considéré que c'était intéressant d'ajouter cette information (même si après il est décidé de ne pas la représenter dans le rendu).

Il faudra l'encoder à la manière d'EpiDoc, dans la balise `<damage>` :

```
<damage>(texte endommagé)</damage>
```

Sauts de ligne, espaces et ponctuation

Sauts de ligne

Figure 4 - *CIFM*, 25, 65, détail.
Cliché CESCM/CIFM

Les sauts de ligne du texte originel devront être représentés avec l'élément `<lb/>` (« line-break »), qui sert à marquer le début d'une ligne. Il faudra donc toujours l'encoder au début de chaque ligne, et non pas à la fin. C'est un élément vide, c'est-à-dire, qui n'accepte pas de contenu à l'intérieur.

Si le saut de ligne apparaît au milieu d'un mot, il faudra y ajouter l'attribut `@break` avec la valeur `'no'`, pour indiquer que c'est effectivement le même mot qui continue à la ligne suivante.

Nous avons décidé d'utiliser l'attribut `@n` dans `<lb>` pour numéroter les sauts de ligne. Pour les inscriptions qui se divisent matériellement en plusieurs parties, il faudra répartir à zéro pour chaque unité graphique (il s'agit d'un usage qui vient de l'édition papier).

Ainsi, pour la figure 4, la numérotation des lignes devrait être indépendante pour les parties supérieure et inférieure du texte.

Espaces

Pendant les premiers siècles du Moyen Âge, les espaces entre les mots n'étaient pas du tout une habitude. Surtout pendant le haut Moyen-Âge, beaucoup de textes sont encore en *scriptio continua* (figure 5), et, lorsque les premiers espaces commencent à apparaître, ils ne sont pas forcément insérés entre les mots de manière régulière. Par exemple, dans la figure 6 (une inscription du XIe – XIIe siècle), pour les mots « motus pietate » il y a plus d'espace entre le « u » et le « s » de « motus »

qu'entre ce même « s » final de « motus » et le « p » initial de « pietate ».

Figure 5 - *CIFM*, HS II, 38. *Cliché CESCM / CIFM*

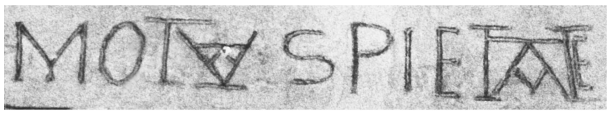

Figure 6 - *CIFM*, 25, 11, l.1. *Cliché CESCM / CIFM*

Il s'agit, sans aucun doute, d'une caractéristique remarquable de l'écriture médiévale, et d'autant plus pour l'épigraphie, car la variabilité des supports, ainsi que l'interaction avec la dimension physique de l'objet et avec l'iconographie, peuvent accentuer la présence d'espaces inattendus dans le déroulement de l'écriture.

Cependant, les espaces physiques entre les lettres ne sont pas représentés dans l'édition papier du *CIFM*, même si souvent leur emploi est décrit en lignes générales dans la description paléographique. Ce n'est pas parce que c'est considéré comme un aspect banal, mais parce que la reproduction fidèle des espaces pose des difficultés à la lecture, et peut empêcher l'accès au texte pour les lecteurs peu habitués à la documentation épigraphique. Comme ici l'édition n'est pas conditionnée par le rendu, il n'y a aucune raison pour ne pas traiter les espaces. On va l'introduire dans le texte d'une manière plus systématique.

Tout d'abord, il faudra distinguer les textes qui font un usage « moderne » des espaces (c'est-à-dire, dans lesquels les mots sont séparés régulièrement par des espaces) de ceux où l'usage des espaces n'est pas normalisé. Pour les premiers, il suffira d'ajouter un espace avec le clavier entre les mots, sans besoin de l'encoder (car il n'y a rien de remarquable dans cet usage, et, par ailleurs, la balise `<space>` pensée pour la description des espaces est réservée uniquement aux espaces qui sont significatifs et inhabituels, et non pas aux espaces « normaux » entre les mots).

Pour les deuxièmes, nous avons décidé de les distinguer des premiers déjà dans la balise `<div type='edition'>`, à travers de l'attribut `@rend='no-space'`, qui va servir à indiquer que, pour l'inscription en question, l'utilisation des espaces n'est pas régularisée, et les mots ne sont pas systématiquement séparés par un espace.

Ensuite, pour chaque espace irrégulier présent dans l'inscription, nous utiliserons l'élément vide `<space/>` pour le signaler. Mais, comme tous les espaces ne sont pas de la même taille, nous avons pensé à les mesurer en fonction de la quantité de caractères du texte qui pourraient rentrer dans l'espace en question. Cela va être encodé à travers les attributs `@unit` (avec la valeur `'character'`) et `@extent` (à remplir avec une valeur numérique équivalente au nombre de lettres qui rentreraient dans l'espace).

Même si tous les caractères n'ont pas la même taille, la mesure de référence approximative pour chaque inscription sera la largeur moyenne des caractères de l'inscription.

La figure 7 contient un fragment d'une inscription comportant un espace significatif au milieu du mot « extitit », entre le « x » et le « t ». Grâce à un montage, dans l'image inférieure de la figure nous pouvons voir une reproduction du même fragment, mais nous avons rempli l'espace existant par une copie exacte du « e » initial du mot (qui nous a paru être de dimensions moyennes), pour montrer que l'espace est équivalent exactement à un caractère, car la lettre remplit parfaitement l'espace manquant. Ainsi, l'encodage de ce fragment par rapport à l'espace va être le suivant :

Figure 7 - *CIFM*, HS II, 52, l.7
Cliché CESCM/CIFM

```
ex<space unit="character" extent="1"/>titit
```

D'autre part, pour la régularisation des espaces dans l'édition normalisée, il ne faudra pas les encoder avec la balise `<space>`. En fait, comme ces espaces suivent des règles très claires (ils apparaissent après chaque mot, sauf s'il est suivi d'un point, une virgule ou un mot enclitique), il suffira de décrire toutes ces informations (d'encoder les mots, la ponctuation et les mots enclitiques) pour pouvoir ensuite les générer automatiquement à partir de la feuille de style.

Interaction avec l'iconographie

Si la succession naturelle du texte est interrompue dans l'inscription non pas par un espace vide, mais par un élément iconographique, il conviendra aussi de le signaler, mais il faudra le traiter différemment de l'espace.

La balise à utiliser dans ce cas-ci sera la balise `<figure>`, et il peut être intéressant de préciser l'élément iconographique causant l'interruption, en écrivant une brève description dans une balise `<desc>` contenue dans `<figure>`.

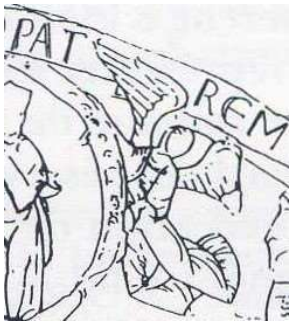

Figure 8 - *CIFM*, 26, 97
Cliché CESCM/CIFM

Par exemple, dans la figure 8 le mot « patrem » est interrompu par l'irruption de l'aile de l'ange qu'il y a sous l'inscription :

```
PATR
<figure>
      <desc>Aile de l'ange de la scène située sous le texte</desc>
</figure>
EM
```

Texte inséré a posteriori

Si l'inscription contient des lettres ou des mots qui n'avaient pas été prévus lors de la conception initiale de l'inscription, et qui ont été rajoutés plus tardivement, il conviendra de le signaler en encodant ce texte dans la balise `<add>`.

Il faudra également préciser, en se servant de l'attribut `@place`, la position de cet ajout par rapport à la ligne d'écriture (une liste de valeurs est proposée par défaut). Pour apporter une description un peu plus détaillée à propos de cette insertion, il faudra se servir de la balise `<desc>`, placée dans `<add>`. Le texte à ajouter devra être placé directement dans la balise `<add>`, à côté de `<desc>`.

Figure 9: CIFM, HS I 28. Cliché CESCM/CIFM.

Par exemple, dans la figure 9, le mot « femina » a été sans doute ajouté plus tard dans l'interligne supérieure :

```
<w>(...)NA</w>
<add>
      <desc>Mot ajouté dans l'espace interligne</desc>
      <w>FEMINA</w>
</add>
<w>QVI</w>
```

Important : il ne faut pas utiliser la balise `<add>` pour signaler les ajouts apportés par l'éditeur de la notice, mais seulement pour des ajouts anciens contenus physiquement dans l'inscription.

ABRÉVIATIONS

Au Moyen-Âge, écrire avec des abréviations devient la façon normale d'écrire. Par rapport à l'époque classique, elles se multiplient et se normalisent, et peuvent apparaître dans tout type de textes. Elles prennent donc une grande importance dans l'étude de l'écriture médiévale, et, par conséquent, dans l'épigraphie médiévale, où les abréviations sont également bien présentes.

Mais, qu'est-ce qu'une abréviation ? C'est le raccourcissement d'un mot ou d'une expression, composé par les lettres présentes qui représentent le mot abrégé, les lettres absentes qui ont été coupées de la représentation et les signes d'abréviation, qui servent à indiquer l'absence de ces lettres. Au Moyen-Âge, il y en a de plusieurs types :

- Selon la manière d'abréger le mot, les abréviations peuvent être par **suspension** (qui comporte la supression de la fin du mot) ou par **contraction** (si le mot conserve le début et la fin, mais quelques lettres du milieu ont été supprimées). Ainsi, par exemple, « ē » en latin est une abréviation par suspension de « est », tandis que « $\overline{\text{DS}}$ » est une abréviation par contraction de « Deus ».
- Selon la manière de signaler l'abréviation, les abréviations peuvent être par **signes spéciaux** (descendants par la plupart des notes tironiennes de l'antiquité classique, ce sont des signes qui indiquent une lettre ou un groupe de lettres concret, comme « 9 » pour « con »), par **lettre suscrite** (si l'abréviation est indiquée par une lettre suscrite du mot abrégé), ou des ***nomina sacra*** (réservés à certains noms chrétiens sacrés, dont l'abréviation s'établit généralement par des caractères grecs).

La manière recommandée d'EpiDoc pour encoder les abréviations a été conçue pour s'adapter aux particularités de l'écriture épigraphique classique et aux intérêts des épigraphistes de cette période (pour lesquels l'analyse paléographique n'était pas une priorité, à cause surtout de la régularité extrême de l'écriture épigraphique classique).

Ce modèle d'encodage prend comme balise racine de l'abréviation l'élément `<expan>`, pensé pour accueillir le mot développé. Le mot n'est écrit qu'une seule fois : la partie abrégée est encodée dans la balise `<abbr>`, et, si elle contient des signes d'abréviation, ils devront rentrer à leur tour dans la balise `<am>` ; et la partie développée est encodée dans la balise `<ex>`.

Voici un exemple d'encodage des abréviations à la manière d'EpiDoc (« $\overline{\text{S}}$ » pour « sanctus ») :

```
<expan><abbr>s<am>¯</am></abbr><ex>anctus</ex></expan>
```

Ainsi, dans la partie abrégée, la balise <am> sert à encoder le signe d'abréviation, qui est à supprimer lors du développement du mot. Mais, qu'est-ce qui se passe avec les abréviations par lettre suscrite, où la lettre appartient à la fois aux lettres de l'abréviation non développée et au signe d'abréviation ? En fait, dans ce cas-ci, le signe d'abréviation à proprement parler, et, par conséquent, ce qui devrait rentrer dans la balise <am>, est seulement la suscription et non pas la lettre en soi, qui est à conserver pour le développement du mot. Mais, avec ce système, cela est impossible à encoder.

Ou, encore plus compliqué, comment faire avec les *nomina sacra*, où les lettres de l'abréviation et les lettres à développer ne sont pas exactement les mêmes ? Cela apparaît complètement impossible à décrire.

Ces exemples (et il y en aurait encore d'autres) servent à montrer que les abréviations ont évolué vers une complexité majeure dans l'époque médiévale, et donc que le modèle d'encodage proposé par EpiDoc est devenu insuffisant.

En fait, les responsables de l'encodage de l'épigraphie médiévale pour le projet ORIFLAMMS avaient constaté ce même problème, et ils avaient opté pour laisser de côté le modèle d'EpiDoc et suivre la manière d'encoder les abréviations proposée par la TEI[7] suivie dans la plupart de projets d'encodage de textes médiévaux, qui consiste à encoder le mot deux fois : d'une part, le mot abrégé est encodé dans la balise <abbr>, tout en signalant les signes d'abréviation avec la balise <am>, ensuite encoder le mot développé dans la balise <expan>, tout en signalant les caractères développés dans la balise <ex>. Les deux parties vont être contenues dans une balise <choice>.

Nous avons décidé de suivre également ce modèle, tout en ajoutant une information supplémentaire qui n'était pas proposée par la TEI ni par ORIFLAMMS, mais qui peut être intéressante : le type d'abréviation.

Cette information va être encodée comme valeur de l'attribut @type de la balise <abbr> (sauf pour les *nomina sacra*, car, comme il s'agit toujours d'abréviations par contraction, nous avons préféré les considérer comme un sous-type des abréviations par contraction, et donc pour les encoder il faudra ajouter, après l'attribut @type='contraction', la catégorie de *nomina sacra* comme valeur de l'attribut @subtype. L'ensemble des valeurs proposées, en fonction du type d'abréviation, est le suivant :

7 Exposée dans le chapitre « 3.3.5. Abbreviations and their expansions » des *TEI P5 Guidelines* (version 3.0.0).

Type d'abréviation	Valeur dans @type	Sous-type d'abréviation	Valeur dans @subtype
Contraction	'contraction'	Nomina sacra	'nomina-sacra'
Suspension	'suspension'		
Contraction et suspension (combinées)	'contraction-suspension'		
Lettres suscrites	'suscription'		
Signes spéciaux	'sign-spec'		

En reprenant l'exemple précédent de « S̄ » pour « s(anctus) », l'encodage serait le suivant :

```
<choice>
        <abbr type="suspension">s<am>-</am></abbr>
        <expan>s<ex>anctus</ex></expan>
</choice>
```

Cette méthode d'encodage résoudrait donc les problèmes avec les abréviations par lettre suscrite et par *nomina sacra*. Puisque l'inscription est encodée deux fois, il suffira juste de décrire dans le premier cas l'abréviation telle qu'elle se trouve dans le texte (avec les caractères spéciaux et les lettres suscrites, s'il y en a), et dans le deuxième cas d'écrire le mot entier, tout en signalant les parties développées avec la balise <ex>.

Par exemple, l'abréviation par suscription « Q^{O} » (pour « q(u)o ») se fera de la manière suivante :

```
<choice>
            <abbr type="contraction" type="suscription">q
                  <am><hi rend="suscription">o</hi></am>
            </abbr>
            <expan>q<ex>u</ex>o</expan>
</choice>
```

Et, pour une abréviation telle que « IHS » pour « Je(su)s », appartenant aux *nomina sacra*, l'encodage serait le suivant :

```
<choice>
        <abbr type="contraction" subtype="nomina-sacra">IHS</abbr>
        <expan>JE<ex>SU</ex>S</expan>
</choice>
```

Usage des entités

Dans l'encodage du projet ORIFLAMMS, les chercheurs avaient donné beaucoup d'importance à l'ergonomie de la saisie des transcriptions et à la réutilisation des données.

Pour atteindre ces objectifs, ils s'étaient servis d'un très bon système dans l'encodage des abréviations : il s'agissait d'encoder uniquement la racine du mot, sans la flexion (car, même si la règle de la TEI est que l'abréviation d'un mot doit être encodée en prenant le mot entier, il est aussi vrai que la flexion n'appartient pas à l'abréviation à proprement parler), et à partir de là construire un système d'entités avec les abréviations.

Ce système permet aux abréviations d'être très facilement réutilisables, ce qui allège la tâche d'encodage. Il aurait été bien de pouvoir se servir du même système et de réutiliser les mêmes entités, mais, les objectifs de TITULUS n'étant pas les mêmes que ceux d'ORIFLAMMS, il y a également de différences substantielles dans les choix d'encodage, car celui de TITULUS est beaucoup plus riche, et vise à donner beaucoup plus d'informations. La différence la plus importante pour l'encodage des abréviations est la représentation des allographes, qui est développée de manière différente dans ORIFLAMMS et dans TITULUS.

Le système des entités est intéressant si l'entité se répète assez souvent et si elle est facile à décrire dans le nom de l'entité (car autrement cela peut prêter à confusion). Dans les premiers essais d'encodage de TITULUS, nous avons essayé de recréer ce même système, mais nous nous sommes rendus compte très vite que, à cause de la richesse de notre encodage, les entités d'ORIFLAMMS n'étaient pas utilisables, et que les nouvelles entités que nous étions en train de créer se multipliaient très vite et se réutilisaient très peu, et que, par ailleurs, il était très difficile à distinguer certaines entités (car, parfois, la seule chose qui changeait était une conjonction qui apparaissait ou une lettre qui devenait onciale).

Bref, la diversité de l'encodage de TITULUS ne s'adaptait pas à l'utilisation des entités pour les abréviations. Ainsi, nous avons laissé de côté cette voie, et nous avons opté pour suivre à la lettre les règles de la TEI que nous venons de décrire.

Forme et apparence

Jeux de lettres

Les jeux de lettres sont un phénomène graphique qui se produit entre deux ou plus caractères qui se combinent par enclavement, entrelacement ou conjonction. Ils sont propres à l'épigraphie médiévale : ils n'existent pas avec cette ampleur et cette recherche d'ordre esthétique pour l'épigraphie classique, et, par conséquent, rien n'est prévu pour leur encodage dans EpiDoc ni dans les normes d'usage de la TEI.

Dans l'édition papier du *CIFM*, les jeux de lettres ne sont pas représentés dans l'édition du texte, mais ils sont toujours signalés dans la description paléographique et dans les index (même si dans ce dernier cas le type n'est jamais spécifié). La raison de leur absence de l'édition en est, encore une fois, la difficulté de les représenter.

Une première tentative d'encodage a été faite dans ORIFLAMMS, ce qui constitue un très bon point de départ. Cependant, pour TITULUS nous en avons profité pour réviser et perfectionner leur modèle, surtout par rapport à la description de la complexité des jeux de lettres (principalement pour les enclavements).

Enclavements et semi-enclavements

En suivant la définition de la professeur Cécile Treffort, un **enclavement** est un « procédé graphique visant à placer (enclaver) une petite lettre à l'intérieur d'une plus grande »[8]. Un **semi-enclavement** est le même type de procédé, mais dans ce cas-ci la lettre d'accueil ne renferme pas complètement la lettre enclavée.

Figure 10 - *CIFM*, HSII, 30, l.2. « M » enclavé par « O ». *Cliché CESCM / CIFM*

La solution adoptée par ORIFLAMMS, qui consistait en encoder les enclavements en signalant la lettre enclavée avec la balise `<hi>` et l'attribut `@rend` (avec la valeur `'enclavé'` ou `'semi-enclavé'`), paraît a priori assez simple et convaincante :

```
<hi rend='enclavé'>(lettre enclavée)</hi>
<hi rend='semi-enclavé'>(lettre semi-enclavée)</hi>
```

8 TREFFORT (Cécile), *Paroles inscrites,* « Glossaire », p.197.

Par exemple, l'encodage de la figure 10 selon le modèle d'ORIFLAMMS se ferait ainsi :

```
DO<hi rend='enclavé'>M</hi>NO
```

Ce modèle d'encodage paraît assez clair, car normalement les enclavements ne concernent que deux lettres, et le caractère enclavé ou semi-enclavé se trouve dans la lettre antérieure du mot. Cependant, même si cela est le plus habituel, la réalité peut être beaucoup plus complexe : l'ordre de succession des lettres peut varier, et la lettre enclavée peut se trouver aussi dans la lettre suivante du mot. Dans ces cas, le modèle d'ORIFLAMMS, qui spécifie la lettre enclavée, mais pas la lettre enclavante, commence à poser un problème.

Par exemple, dans la figure 11, pour le fragment « iulii obiit », le dernier « i » de « iulii » est enclavé dans le « o » initial d' « obiit ». Ici, la manière d'encoder d'ORIFLAMMS (« `iuli<hi rend='enclavé'>i</hi> obiit` ») créerait une confusion, car, même s'il paraît évident que le « i » enclavé ne peut pas être dans le « i » antérieur, le « o » enclavant est en fait la lettre suivante et non pas celle précédente, et par ailleurs elle appartient à un nouveau mot.

Figure 11 - *CIFM*, 26, 157, pourtour du linge. *Cliché CIFM/CESCM*

D'autre part, il est également possible de trouver des enclavements concernant plus de deux lettres. Dans la figure 12, par exemple, nous avons deux images d'enclavements de trois lettres, mais qui présentent une différence assez remarquable entre eux : tandis que le premier est un « A » qui enclave les lettres « T » et « I », qui se situent côte à côte, dans le deuxième il y a un « P » qui enclave un « A » qui, à son tour, enclave un « R ». Dans le premier exemple nous avons une lettre qui enclave deux autres lettres, et dans le deuxième exemple nous avons une lettre qui enclave une lettre qui enclave une lettre (c'est-à-dire, un enclavement dans un enclavement).

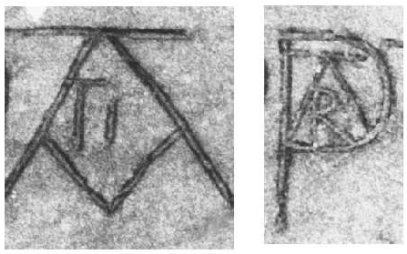

Figure 12 - *CIFM*, 25, 11, détails. *Cliché CIFM/CESCM*

Ici encore une fois, le modèle d'encodage d'ORIFLAMMS ne ferait pas de distinction entre les deux enclavements (et, d'autre part, il n'y a rien qui puisse nous indiquer que les deux lettres enclavées dans chaque cas appartiennent en fait au même enclavement) :

```
A<hi rend='enclavé'>T</hi><hi rend='enclavé'>I</hi>
P<hi rend='enclavé'>A</hi><hi rend='enclavé'>R</hi>
```

Pour le premier, nous pourrions encoder le « T » et le « I » dans la même balise (`A<hi rend='enclavé'>TI</hi>`), ce qui servirait peut-être à faire une première distinction, mais même comme ça, l'encodage ne décrit pas assez clairement le phénomène.

La raison de ce choix par ORIFLAMMS est, en fait, très intéressante et très innovatrice, car ce projet avait proposé de redéfinir la conception des enclavements pour les considérer comme « des lettres entrant en combinaison avant et arrière ». Cependant, pour l'instant seule la combinaison vers l'avant est prévue par Unicode et Mufi, et il n'y a aucune manière d'indiquer la combinaison vers l'arrière. La proposition d'ORIFLAMMS est malheureusement trop avancée, et il faudra se contenter de trouver une solution peut-être moins innovatrice, mais plus précise dans la description. Pour ce faire, il s'agira d'encoder aussi les caractères « enclavants », et non seulement les enclavés.

A priori, il y aurait deux manières différentes d'encoder la balise `<hi rend='enclavant'>` : y faire rentrer la lettre enclavante et la balise avec la lettre enclavée, ou n'y faire rentrer que la lettre enclavée. La première option est sans doute la plus optimisée, car elle nous définit les limites de l'enclavement. Cependant, elle n'est pas très correcte, car on introduit la partie enclavée dans la partie définie comme enclavante, et donc la lettre enclavée serait définie aussi par cette balise. La seconde option est plus correcte, car les informations y apparaissent bien séparées. L'ordre des lettres lui aussi est beaucoup plus clair. Cependant, cette solution d'encodage apporte moins d'information, car l'enclavement n'est pas délimité.

Nous avons choisi l'option 2, et nous avons ajouté une balise `<seg>` (neutre pour définir un segment de texte quelconque) qui englobe les deux balises d'enclavant et enclavé.

Il nous a paru plus adéquat d'utiliser la balise `<seg>` que `<hi>` pour délimiter l'enclavement, car `<hi>` « distingue un mot ou une expression comme graphiquement distincte du texte environnant, sans en donner la raison » (selon les *TEI Guidelines*), tandis que `<seg>` sert à signaler un segment de texte d'une manière assez neutre. Et, tandis que le balisage des lettres enclavantes et enclavées vise à les rendre graphiquement distinctes, la balise <seg> sert tout simplement à établir les limites de l'enclavement.

Le modèle sera donc le suivant :

```
<seg rend='enclavement'>
            <hi rend='enclavant'> (lettre enclavante) </hi>
```

```
                <hi rend='enclavé'> (lettre enclavée) </hi>
    </seg>
```

Si l'ordre des lettres est inversé (c'est-à-dire, que la lettre enclavée précède la lettre enclavante dans l'ordre naturel de la phrase), cela ne poserait aucun problème pour l'encodage :

```
    <seg rend='enclavement'>
                <hi rend='enclavé'>(lettre enclavée)</hi>
                <hi rend='enclavant'>(lettre enclavante)</hi>
    </seg>
```

Pour une lettre qui enclave deux lettres qui sont toutes les deux au même niveau, il suffira d'inclure les deux lettres enclavées dans la même balise `<hi rend='enclavé'>` :

```
    <seg rend='enclavement'>
                <hi rend='enclavant'>(lettre enclavante)</hi>
                <hi rend='enclavé'>(lettres enclavées)</hi>
    </seg>
```

Et pour une lettre qui enclave une lettre qui enclave une autre lettre, la solution sera d'insérer, dans la balise `<hi rend='enclavé'>` du premier niveau, la lettre enclavée par la première lettre mais qui à son tour en enclave une autre dans une balise `<hi rend='enclavant>`, et la lettre doublement enclavée dans la balise `<hi rend='enclavé'>` :

```
    <seg rend="enclavement">
          <hi rend="enclavant">(lettre qui enclave un enclavement)</hi>
          <hi rend="enclavé">
                <hi rend="enclavant">(lettre enclavée par le première
                lettre et enclavant la troisième)</hi>
                <hi rend="enclavé">(lettre doublement enclavée)</hi>
          </hi>
    </seg>
```

D'autre part, pour les semi-enclavements, en principe la procédure d'encodage devrait être presque la même, car la seule chose qui change est le résultat final (puisque la lettre enclavante n'enferme pas complètement la lettre enclavée). Nous avons décidé d'utiliser exactement le même procéder, et de changer seulement la dénomination du `<seg>` (`<seg rend="semi-enclavement">`), tout en conservant `<hi rend="enclavé">` et `<hi rend="enclavant">` pour l'encodage des lettres.

Finalement, si, comme nous avons vu dans la figure 11, l'enclavement se déroule entre deux

mots, il faudra encore une fois privilégier l'unité du mot. Par conséquent, le balisage de l'enclavement va être divisé entre les deux mots, tout en les connectant à travers d'un attribut `@xml:id` dans le premier mot et un attribut `@corresp` dans le deuxième qui renvoie au contenu du premier :

```
<w>…
      <seg rend="enclavement" xml:id="enc-julii-obiit">
            <hi rend="enclavé">I</hi>
      </seg>
</w>
<w>
      <seg rend="enclavement" corresp="#enc-julii-obiit">
            <hi rend="enclavant">O</hi>
      </seg>…
</w>
```

Entrelacement

En nous remettant toujours aux descriptions des jeux de lettres faites par Cécile Treffort, un entrelacement est un « procédé graphique consistant à réunir plusieurs lettres en traçant les traits de manière à ce qu'ils se recoupent sans se confondre. »[9]

Comme pour l'enclavement, dans ORIFLAMMS était décrite seulement une des deux lettres entrelacées :

```
<hi rend='entrelacé'>(lettre entrelacée)</hi>
```

Mais, est-ce qu'il y a une relation de subordination entre les deux lettres entrelacées ? En fait, selon la définition de l'entrelacement, ce sont les deux lettres qui sont entrelacées, dans une situation de réciprocité, et non pas une qui est entrelacée par l'autre. En suivant cette définition, donc, il ne paraît pas très justifié d'encoder seulement une des lettres entrelacées et pas l'autre. Cependant, il est aussi vrai que souvent l'entrelacement ressemble beaucoup à l'enclavement (avec une lettre qui est plus grande horizontalement et verticalement, et une lettre plus petite qui se trouve au milieu de la première), avec une relation de subordination très évidente, et la seule chose qui change par rapport à l'enclavement est que les traits des deux lettres se coupent entre eux. C'est le cas de la première image de la figure 13 avec le jeu de lettres entre le « E » et le « M », où, sauf pour le trait central du « E », l'entrelacement a toutes les caractéristiques d'un semi-enclavement.

9 TREFFORT (Cécile), *Paroles inscrites*, « Glossaire », p. 197.

Figure 13 - *CIFM*, 20, 28 ; *CIFM*, 21, 191 ; *CIFM*, 23, 104.
Détails d'entrelacements. *Cliché CESCM/CIFM*

En revanche, il y a d'autres cas où le rapport entre les lettres est beaucoup plus équilibré, comme dans la deuxième et la troisième image de la figure 13, où les lettres entrelacées ont la même hauteur et, dans la deuxième image, aussi le même longueur, et il est impossible d'établir une hiérarchie claire.

Des réflexions avaient été faites pour ORIFLAMMS, et il y avait eu même des propositions pour mettre au même niveau tous les jeux de lettres.

Mais cela n'a pas abouti, et finalement les distinctions ont été gardées. Pour TITULUS, nous avons décidé de maintenir les mêmes précautions, car, même si la frontière entre les termes n'est pas claire, on a besoin d'une systématisation pour pouvoir les traiter d'une manière efficace (non seulement par la machine, mais aussi par les chercheurs). D'autre part, il ne faut pas oublier qu'il s'agit de termes paléographiques bien figés et qui ont une acceptation incontestable dans l'étude de l'épigraphie et la paléographie médiévales. Il faudrait des arguments plus solides et incontestables pour y introduire des modifications.

En tout cas, en conservant cette distinction dans l'encodage, il sera toujours possible de traiter informatiquement comme un même groupe les enclavements et les entrelacements s'il nous paraît pertinent. Par contre, le contraire ne serait pas possible et requerrait un travail supplémentaire d'encodage.

Ainsi, l'encodage de l'entrelacement va être maintenu pour TITULUS, et le mieux sera de considérer que les lettres qui le forment se trouvent au même niveau (car, comme nous l'avons vu, il n'est pas toujours possible d'établir une relation de subordination). De cette manière, il ne faudra pas les encoder séparément pour distinguer la fonction de chacune d'entre elles, mais il suffira de les encoder ensemble dans la même balise (`<seg @rend='entrelacement'>`) avec l'ordre de succession du mot.

Conjonction

Une conjonction est (toujours selon Cécile Treffort) un « procédé graphique fondé sur la réunion de deux lettres, voire plus, par fusion de deux hastes ou d'une haste et d'un trait courbe »[10].

Figure 14 - *CIFM*, HS II, 2 ; *CIFM*, HS II, 52 ; *CIFM*, HS II, 49 ; *CIFM*, HS II, 51. Conjonctions « AV », « TE », « NE », « AP » et « LV ».
Cliché CESCM/CIFM

Les conjonctions sont le jeu de lettres le plus commun. Elles ne sont pas exclusives de l'épigraphie médiévale, et en fait leur encodage est déjà prévu dans la TEI par la balise `<hi rend='ligature'>`. C'est justement cette formule qu'ORIFLAMMS a reprise pour les encoder.

Comme pour l'entrelacement, il faut considérer que les deux lettres conjointes sont au même niveau, et donc qu'il faut les traiter de la même manière.

Figure 15 - HS II, détails.
Cliché CESCM/CIFM

D'autre part, il faut faire attention, car toutes les lettres qui partagent des hastes ne sont pas considérées comme des conjonctions à proprement parler par les épigraphistes médiévistes. Un bon exemple en est l'écriture gothique minuscule du XVe siècle (figure 15), dans laquelle le trait vertical du « e », lorsque la lettre est précédée d'un « p », d'un « b » ou d'un « d », apparaît unifié avec la partie droite de la lettre antérieure. Cela est considéré comme une caractéristique propre au type d'écriture, qui est loin de l'esprit des jeux de lettres de l'épigraphie médiévale.

Ainsi, pour ne pas avoir de confusion avec une ligature quelconque qui n'est pas un jeu de lettres, et pour maintenir la même homogénéité avec les autres jeux de lettres, nous avons décidé de ne pas encoder les conjonctions sous `<hi rend='ligature'>`, mais sous `<seg rend='conjonction'>`. De cette manière, il serait possible d'encoder à la fois la ligature du « e » de l'écriture gothique comme une ligature quelconque (`<hi rend='ligature'>`), et d'encoder aussi d'une manière distincte la ligature propre aux jeux de lettres (`seg rend='conjonction'`).

10 TREFFORT (Cécile), *Paroles inscrites*, « Glossaire », p. 196.

Combinaisons de plusieurs jeux de lettres différents

Parfois, les jeux de lettres apparaissent en combinaison avec d'autres jeux de lettres, ce qui rend l'encodage plus complexe.

Le plus fréquent est que la combinaison entre deux jeux de lettres différents se fait entre une conjonction et un enclavement et une conjonction et un entrelacement. Dans ce cas-là, la meilleure solution pour l'encodage d'un jeu de lettres multiple sera de considérer l'entrelacement ou l'enclavement comme une partie de la conjonction. De cette manière, il n'y aura pas de difficultés pour encoder des jeux de lettres tels que celui de la figure 16, où la conjonction se produit entre deux lettres qui ne sont pas successives dans l'ordre de la phrase :

Figure 16 - *CIFM* HS II, 34. Entrel.« NI » et conj. « NG ». *Cliché CESCM / CIFM*

```
    BE
<seg rend="conjonction">
                <seg rend="entrelacement">NI</seg>G
    </seg>NUS
```

Figure 17 - *CIFM*, 7, 49. *Cliché CESCM/CIFM*

Cette solution s'avère être utile aussi pour des combinaisons plus complexes telles que celle de la figure 17, où le jeu de lettres est formé par quatre lettres, dont les deux premières et la quatrième (dans l'ordre du mot) conforment une triple conjonction, et au même temps la troisième est enclavée par la quatrième :

```
                    <seg rend="conjonction">qu
                <seg rend="enclavement">
                     <hi rend="enclavé">o</hi>
                     <hi rend="enclavant">n</hi>
                </seg>
    </seg>
```

Il est vrai que ce système peut sembler peu précis, car on ne spécifie pas avec laquelle des deux lettres de l'entrelacement ou de l'enclavement se produit la conjonction. Dans le cas de l'enclavement, en principe cela devrait être toujours avec la lettre enclavante, mais pour l'entrelacement les deux lettres se trouvent au même niveau, et donc il est impossible de savoir avec quelle lettre se produit la

conjonction. Cependant, compte tenu de la complexité du phénomène graphique à représenter, cette solution nous paraît la plus adéquate.

Figure 18 - *CIFM*, 7, 49.
Cliché CESCM/CIFM

D'autre part, même si c'est plus rare, il n'est pas impossible de trouver des jeux de lettres complexes comportant la combinaison d'entrelacements et d'enclavements. Dans la figure 18, par exemple, le mot « QVID » est composé par un double entrelacement des lettres « Q », « V » et « D », et à son tour la lettre « Q » enclave la lettre « I ». Par ailleurs, l'enclavement se produit entre deux lettres qui ne se succèdent pas dans le mot, ce qui complique encore plus l'encodage.

La solution que nous avons trouvée pour décrire ce mot est de considérer tout le mot comme un entrelacement, et ensuite d'y faire rentrer l'enclavement, divisé en deux parties liées par un identifiant commun, afin de maintenir l'ordre de succession des lettres dans la phrase :

```
<seg rend="entrelacement">
      <seg rend="enclavement" xml:id="encl-q-i">
            <hi rend="enclavant">Q</hi>
      </seg>
      V
      <seg rend="enclavement" corresp="#encl-q-i">
            <hi rend="enclavé">I</hi>
      </seg>
      D
</seg>
```

Caractères et glyphes

La tradition du *CIFM*, en suivant le modèle d'édition de l'épigraphie classique, était de transcrire toutes les inscriptions en capitales, même si elles ne l'étaient pas. D'autre part, les caractères à formes spéciales n'étaient pas signalés dans la transcription à cause de la difficulté de les représenter. Ils n'étaient pas décrits dans ORIFLAMMS non plus.

Cependant, la tradition épigraphique médiévale a réfléchi depuis longtemps aux formes des lettres des inscriptions, car elles sont d'une grande richesse et singularité par rapport aux autres types d'écriture médiévaux. L'école poitevine n'est pas une exception, mais dans le *CIFM* cette information est réservée au commentaire paléographique.

Il s'agira donc ici de mener une réflexion pour les représenter d'une manière systématique. Pour ce faire, le mieux est de commencer par poser les bases du vocabulaire, à partir des réflexions de deux des plus grands spécialistes dans le traitement digital de la paléographie médiévale, les chercheurs Peter Stokes (King's College) et Dominique Stutzmann (IRHT).

Peter Stokes, dans le projet DIGIPAL (un projet mené par le King's College de Londres visant à fournir des outils informatiques pour le traitement et l'étude de la paléographie médiévale), avait cru indispensable la définition d'un vocabulaire solide et stable des éléments composant l'écriture médiévale comme base pour fonder son projet[11]. Ce vocabulaire a été traduit et révisé par Dominique Stutzmann dans son article « *Modélisation des signes graphiques* »[12].

Sans trop entrer dans le détail, pour ce qui nous intéresse dans l'étude des formes des lettres dans l'épigraphie médiévale il faut distinguer entre les catégories suivantes :

Graphème, **lettre** ou **signe :** Unité linguistique minime avec valeur sémantique. Dans notre cas, il s'agit des composants de l'alphabet latin médiéval (qui est à peu près le même que l'actuel, mais sans distinguer le U du V ni le J du I).

Allographe: chacune des formes idéelles qu'un même graphème, lettre ou signe peut acquérir.

Figure 19
Cliché CESCM/CIFM

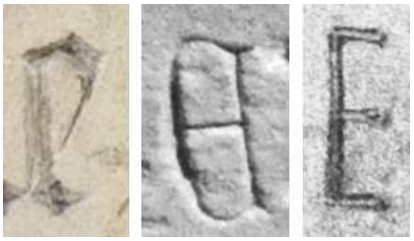

Figure 20 - *CIFM*, HS II ; 26, 102 ; 25, 11. *Cliché CESCM/CIFM*

Ainsi, par exemple, la figure 20 présente une même lettre (« E ») représentée par trois allographes distincts (un « e » minuscule, un « e » oncial et un « e » capital), tandis que dans la figure 19 on peut parler en fait d'un même allographe, le « e » oncial, même s'il est exécuté par des auteurs différents, dans des supports différents et avec quelques variantes dans l'exécution.

Idiographe : prototype, manière dont un certain scribe réalise un allographe. Par exemple, la

11 Cette terminologie est exposée dans le glossaire du site web du projet : http://www.digipal.eu/help/glossary/

12 STUTZMANN (Dominique), « Modélisation des signes graphiques (1) », *Paléographie médiévale*, 4 février 2012. URL : https://ephepaleographie.wordpress.com/2012/02/04/modelisation-des-signes-graphiques-1/. Consulté le 10 août 2016.

figure 19 contient trois idiographes différents du « e » oncial, tandis que la figure suivante contient quatre représentations différentes du même idiographe (car elles ont été réalisées par le même auteur) :

Figure 21 - *CIFM*, 26, 102. *Cliché CESCM/CIFM*

Graphie : une représentation concrète d'un graphème déterminé. (Ainsi, chacun des allographes de la figure 20, et chacun des idiographes de la figure 19 serait une graphie unique et distincte par rapport au reste).

Cela est pour la considération des lettres dans la paléographie médiévale. Mais il sera intéressant de voir aussi la correspondance des termes dans Unicode (le standard international d'encodage des caractères en informatique). Les deux principales distinctions sont les suivantes :

Caractère : « le composant le plus petit d'un langage avec valeur sémantique »[13], ce qui se correspondrait à la lettre déliée de sa forme.

Glyphe : forme spécifique que ces caractères peuvent prendre (qui serait l'équivalent d'un allographe).

À partir de toutes ces définitions, en reprenant les termes que nous venons de voir, ce qui nous intéresse et qui a raison d'être signalé dans la transcription sont les allographes (des modèles formels qui peuvent se répéter dans des inscriptions différentes), et ils devront être encodés, selon la terminologie d'Unicode, comme des glyphes (qui devront être mis en rapport à chaque fois avec le caractère correspondant).

Avant de passer à traiter leur encodage, il sera utile d'approfondir la réflexion sur la nature des allographes et la typologie des formes qu'ils peuvent adopter :

Minuscules, onciales, capitales

Il faut préciser surtout (car même s'il paraît évident, la question peut se poser avec l'immense variété des inscriptions médiévales) qu'il ne s'agit pas ici de décrire la taille, mais la forme de la lettre. Dans les fonts modernes, les minuscules sont systématiquement plus petites que les majuscules, mais

13 *Glossary of Unicode Terms*, http://unicode.org/glossary/index.html.

au Moyen Âge cela n'est dans aucun cas une règle, et encore moins pour l'épigraphie médiévale (où la variabilité de l'écriture est très importante).

D'autre part, à côté de la minuscule et la capitale il faut considérer la graphie **onciale**, qui se situe à mi-chemin entre les formes capitales et minuscules, et qui est assez fréquente dans les inscriptions médiévales.

Dans les textes épigraphiques médiévaux, lorsqu'on trouve des mélanges de caractères onciaux, minuscules et / ou majuscules dans la même inscription (souvent sans que cela soit imposé par un critère orthographique ou sémantique), on voit que la distinction se fait par la forme, mais que cela n'affecte pas toujours la taille de la lettre. Ainsi, par exemple, dans la figure 22, dans un mot où toutes les lettres sont de la même taille (et majoritairement en capitales), le premier « E » est capital mais les deux derniers sont onciaux.

Figure 22 - *CIFM* 25, 26. Détail du texte *Cliché CESCM/CIFM*

D'autre part, il ne s'agira évidemment pas d'encoder tous les caractères du texte, car cela serait extrêmement lourd et non nécessaire. Dans l'immense majorité des textes épigraphiques du *CIFM*, dans l'écriture du texte il y a toujours une casse qui domine (normalement des capitales), qui est alternée avec certains caractères dans une autre casse. Il s'agira donc de signaler seulement les caractères qui ne sont pas majoritaires.

Ainsi, prenons par exemple la figure suivante, dans laquelle la majorité des caractères sont en capitale, mais quelques-uns sont onciaux.

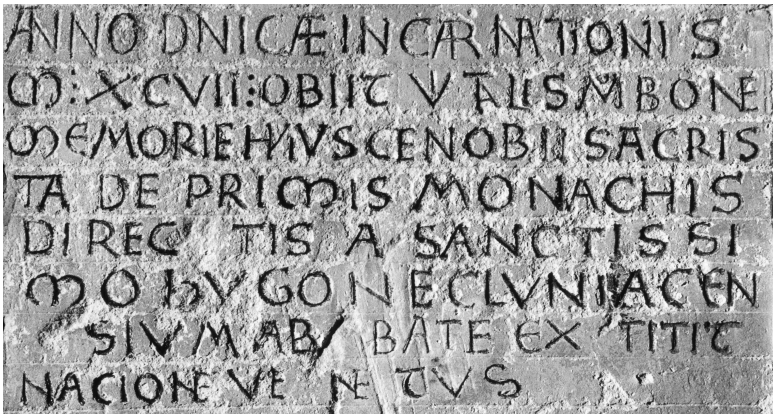

Figure 23 - *CIFM*, HS II, 52. *Cliché CESCM / CIFM*

Ce qu'il faudra faire sera de repérer les caractères onciaux :

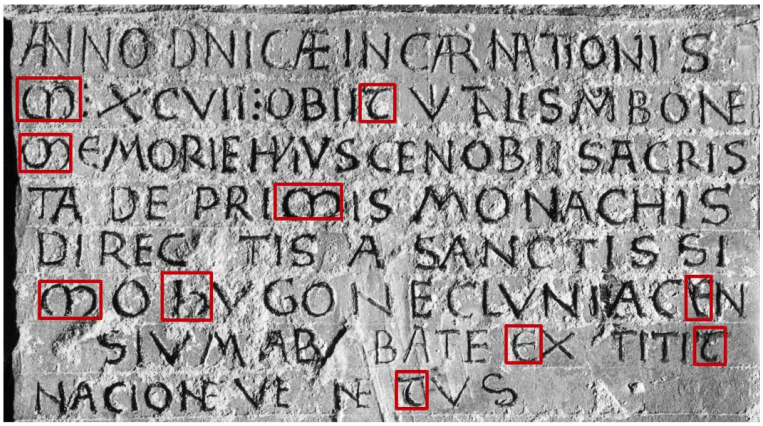

Figure 24 - *CIFM*, HS II, 52. *Cliché CESCM / CIFM*

Finalement, dans l'encodage, il s'agira de tout transcrire en capitales, sauf pour les caractères onciaux, qui vont être encodés comme tels.

D'autres allographes

Mais les allographes ou les glyphes ne sont pas que les minuscules, les capitales et les onciales. En fait, un caractère dans une même casse peut être représenté par des allographes différents, comme dans la figure 25 avec le « A » capital :

Figure 25 - *CIFM*, 25, 11 ; 25, 68 ; 26, 102.
Cliché CESCM/CIFM

Dans ces trois cas, qui sont des extraits d'inscriptions appartenant au *CIFM*, l'identification distincte de la forme des lettres a été mise en relief dans la description paléographique par les éditeurs. Mais, où établir les limites de séparation et de convergence entre les glyphes ? Où termine un graphe ou idiographe et commence un allographe ? L'allographe étant une forme idéelle, il faut systématiser pour reconnaître dans chaque idiographe l'allographe qu'il a comme modèle.

Dans ce sens, l'épigraphie médiévale comporte une difficulté ajoutée (qui apparaît beaucoup moins dans l'écriture médiévale sur papier ou parchemin) : parfois, un caractère adopte des formes qui peuvent nous paraître remarquables du point de vue paléographique, mais qui en fait sont conditionnées par le support, ou par la manière dont les lettres ont été stylisées et décorées, et il ne s'agit pas d'une question particulière concernant la forme de la lettre.

Figure 26 - *CIFM*, 26, 157. Partie supérieure. *Cliché CESCM/CIFM*

Par exemple, dans la figure 26, même si l'exécution de l'écriture est tout à fait particulière, il ne faut pas attribuer cette singularité à la forme propre à chaque lettre, mais au style de l'ensemble de l'exécution, qui est externe à la réalisation allographétique.

Un allographe est une variante graphique d'une lettre qui affecte significativement sa morphologie. Quelques exemples typiques de l'épigraphie médiévale sont, par exemple, le « O » en losange, le « C » carré ou le « A » avec une barre transversale brisée (figure 27).

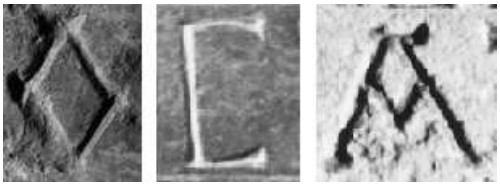

Figure 27 - *CIFM*, HS II, 2 ; HS II, 31 ; HS II, 69. *Cliché CESCM/CIFM*

Pour l'encodage de TITULUS, le plus sage sera d'identifier comme des allographes seulement les variantes formelles de caractères qui apparaissent décrites dans le commentaire paléographique des notices.

Transcription

Comme la correspondance entre les glyphes sur Unicode et les allographes existants pour l'épigraphie médiévale est loin d'être satisfaisante (car, d'un côté, Unicode ne reconnaît pas les différents glyphes comme appartenant à un même caractère, et de l'autre côté tous les allographes définis par l'épigraphie médiévale ne se trouvent pas dans Unicode), avant d'encoder directement la lettre dans le texte, il faudra faire un travail préalable, afin de donner une description du glyphe et le

mettre en rapport avec le caractère correspondant.

Cela est déjà prévu dans la TEI et EpiDoc. Il faut encoder cette information dans `<charDecl>` (dans `<encodingDesc>`, comme nous l'avons déjà signalé), dans une balise `<glyph>` pour chaque allographe.

Il faudra une balise `<glyph>` pour chaque allographe à décrire. Chaque balise portera un attribut identifiant (`@xml:id`), qui aura pour valeur la lettre, un tiret et une description très synthétisée du glyphe (par exemple, pour le s long, ce sera « `s-long` », ou pour le e minuscule, « `e-minuscule` »). Dedans, il devra y avoir toute l'information concernant le caractère, organisée en balises :

- Une balise `<glyphName>` contenant une description succincte du glyphe (par exemple « A sans traverse horizontale », « O en losange » ou « T oncial »).
- Une balise `<mapping>` comportant la forme la plus similaire possible au glyphe dans sa forme idéale. Si le caractère appartient aux caractères standard de l'alphabet latin (regroupés dans la norme ISO 646, connu aussi comme « Latin Basique » ou « Alphabet International nº 5 »), il suffira de l'écrire comme tel dans le contenu de la balise. Dans ce cas-ci, il faudra ajouter l'attribut `@type='standard'`, pour signaler qu'il s'agit du caractère standardisé auquel le glyphe fait référence.

  Si, par contre, le caractère n'est pas un caractère latin standard, il faudra aller voir s'il existe sur Unicode, et l'encoder avec le code alphanumérique correspondant, précédé de `'&'` et terminé par `';'`. Dans ce cas-là, l'attribut à ajouter à la balise `<mapping>` sera `@type='unicode'`. Ensuite, il faudra ajouter une balise `<mapping>` supplémentaire, comportant l'attribut `@type='standard'`, et contenant la forme régularisée du glyphe. Il s'agit ici de faire le lien entre le glyphe et le caractère correspondant.

  Par exemple, pour le « R rond », « 2 » :

```
<glyph xml:id="R-rond">
        <glyphName>R rond</glyphName>
        <mapping type="standard">R</mapping>
        <mapping type="unicode">&#UA75A;</mapping>
</glyph>
```

Si le caractère ne se retrouve pas dans Unicode, il y aura encore une option pour le visualiser, grâce à MUFI (Medieval Unicode Font Initiative), un projet qui cherche à encoder et à rendre

visibles des caractères spéciaux des textes médiévaux écrits en alphabet latin et qui n'appartiennent pas (encore) à Unicode. La liste de caractères de MUFI est disponible sur Internet[14], et il suffira de taper le code du symbole désiré dans la balise `<mapping>`, et d'y insérer l'attribut et la valeur `@type='PUA'` (« PUA » étant les sigles de « Private User Area », qui définit les caractères qui ne sont pas encore dans Unicode). La seule contrainte sera que, pour la visualisation, il faudra choisir une police qui soit compatible avec MUFI.

La seule différence par rapport à Unicode est que, pour visualiser les caractères de MUFI du type PUA, il faut utiliser une police spéciale qui soit compatible avec eux.

```
<glyph xml:id='s-long'>
      <glyphName>s minuscule long</glyphName>
      <mapping type='PUA'>&#x017F;</mapping>
      <mapping type='standard'>s</mapping>
</glyph>
```

Si, après avoir parcouru MUFI, il n'y a toujours pas de représentation possible pour le caractère, on se contentera seulement d'avoir la balise `<mapping type='standard'>` avec la forme standardisée du caractère. Si jamais le caractère est rajouté à Unicode ou à PUA (car la liste des caractères est révisée périodiquement, et de nouveaux caractères sont rajoutés de temps en temps), on pourra rajouter facilement la balise correspondante.

Une fois remplie la balise `<glyph>`, nous aurons toutes les informations nécessaires concernant le glyphe : une description écrite, une représentation graphique et le caractère auquel il correspond.

Pour encoder le caractère dans le texte lorsqu'il apparaît, il faudra se servir de la balise `<g>` (*gaiji*), avec l'attribut `@corresp`, qui va avoir comme valeur l'identifiant du glyphe correspondant.

Nous avons considéré qu'il n'était pas nécessaire de remplir la balise `<g>`, car toutes les informations ont été déjà données dans la balise `<glyph>`.

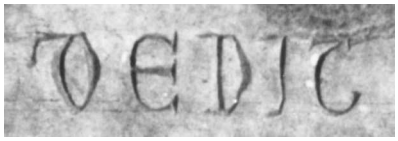

Figure 28 - CIFM, 15, 71.
Cliché CIFM / CESCM

14 *MUFI character recommendations,* version 4.0 : http://folk.uib.no/hnooh/mufi/specs/MUFI-Alphabetic-4-0.pdf

Par exemple, pour transcrire « DEDIT » (figure 28), la transcription va être la suivante :

```
<g corresp='#d-oncial'/><g corresp='#e-oncial'/>DI<g corresp='#t-oncial'/>
```

Usage des entités

Il sera très pratique d'utiliser des entités pour transcrire les balises de description des glyphes, afin de pouvoir les avoir toutes regroupées dans un document externe commun à toutes les notices, et s'assurer que les glyphes décrits et employés sont les mêmes dans l'ensemble du corpus. Par ailleurs, pour les caractères qui ne se retrouvent pas dans MUFI, si jamais on trouve une solution d'affichage qui nous convient mieux (ou que le caractère rentre finalement dans MUFI), le système des entités permettrait de le rajouter très facilement (car on n'aurait qu'un seul document à modifier).

Il sera également utile d'utiliser des entités pour transcrire les allographes dans le texte. Nous avons défini une manière standardisée pour les noms des entités, composées par la lettre, un tiret et un mot définissant le glyphe. Par exemple, pour un « s » dont la forme est renversée, le nom de l'entité sera `'s-renversé'`.

Taille et position des caractères

Parallèlement à la description des allographes, il faudra pouvoir avoir une description des caractères qui ne concerne pas leur forme, mais la position et la taille, et donc qui ne peut pas être encodée dans `<glyph>`.

Et, de la même manière qu'un « e » minuscule suscrit reste un « e » minuscule, cet ordre conceptuel d'idées devra être reproduit dans l'encodage en TEI. L'encodage de ces caractéristiques se fera externellement à celui des allographes, à travers d'une balise `<hi>` avec l'attribut `@rend` et la valeur de la caractéristique correspondante qu'on veut décrire. Il y a déjà une liste de valeurs prédéfinies dans la TEI (qui a été gardée pour EpiDoc), et qui est assez complète.

Ponctuation

La balise réservée à l'encodage de la ponctuation dans la TEI et EpiDoc est `<pc>`, qui a été également utilisée dans ORIFLAMMS. Le balisage de la ponctuation, qui n'est apparu qu'assez tardivement dans la TEI[15], est pourtant très important, spécialement dans les textes anciens, où la

15 En 2008, Alexei Lavrentiev (ENS de Lyon) a montré dans un article intitulé « A tag for punctuation ? » la

ponctuation ne compte pas avec les mêmes signes ni les mêmes usages qu'aujourd'hui, et l'insertion de ces signes sans être encadrés dans une balise qui les définit pourrait prêter à confusion (surtout parce que la plupart d'entre eux ont d'autres signifiés ou bien n'existent pas dans l'écriture actuelle).

La seule chose dont il faudra tenir compte ici par rapport à ORIFLAMMS c'est la distinction entre l'encodage de la ponctuation présente dans l'inscription (à reproduire dans la transcription) et l'encodage de la ponctuation moderne (qui sera à rajouter pour l'édition normalisée). Comme dans ORIFLAMMS il n'y avait que la ponctuation originelle de l'inscription, la question sur la distinction ne s'était pas posée. La meilleure manière sera de les distinguer en les encodant dans les balises `<orig>` (pour la ponctuation médiévale) et `<reg>` (pour la ponctuation moderne).

Ponctuation originelle

Il s'agit de reproduire la ponctuation exactement où elle se trouve dans le texte. Mais, de même que pour l'encodage des lettres, il y a des symboles de ponctuation qui n'existent plus dans la ponctuation moderne, et donc qui peuvent ne pas être dans Unicode. Encore une fois, dans ces cas-ci, il faudra aller chercher dans MUFI, ou, s'ils n'y sont pas, chercher des solutions avec un peu d'imagination (par exemple, pour les trois points verticaux, une solution pour les reproduire pourra être d'insérer un point suscrit sur deux points verticaux).

L'encodage devra se faire, comme déjà signalé, en encodant la balise `<pc>` dans la balise `<orig>` :

```
<orig>
    <pc>[signe de ponctuation médiévale]</pc>
</orig>
```

Ponctuation moderne

Parallèlement à la ponctuation originelle, il faudra encoder également de la ponctuation moderne qui devra apparaître dans l'édition normalisée, en se servant de la balise <reg> (qui sert à indiquer qu'il s'agit d'une régularisation) :

```
<reg>
    <pc>[signe de ponctuation moderne]</pc>
</reg>
```

pertinence de l'encodage de la ponctuation dans la TEI, en signalant sa singularité et son importance comme objet de recherche. Peu de temps après, le consortium de la TEI a introduit la balise <pc> dans le schéma de la TEI.

Si un signe de ponctuation ancien est à conserver pour la ponctuation moderne (ce qui n'est pas très habituel pour l'épigraphie médiévale), il suffira de l'encoder deux fois, une avec la balise `<orig>` et l'autre dans la balise `<reg>`.

Usage des entités

Dans les deux cas, il sera pratique de se servir du système des entités pour l'encodage de la ponctuation dans le texte. La manière de composer les noms des entités, inspirée de celle d'ORIFLAMMS, est la suivante :

- pour les signes de ponctuation anciens, les lettres « pc » (pour « ponctuation ») suivies d'un tiret et du nom du signe de ponctuation. Par exemple, pour le point virgule, l'entité est celle qui suit :

```
<!ENTITY pc-point-virgule "<orig><pc>;</pc></orig>">
```

- pour les signes de ponctuation modernes, le modèle sera exactement le même, mais en substituant tout simplement le préfixe « pc » pour « pcm » (pour « ponctuation moderne »). Voici l'encodage du point virgule dans la ponctuation moderne :

```
<!ENTITY pcm-point-virgule "<reg><pc>;</pc></reg>">
```

Organisation syntaxique et grammaticale

Mots

L'identification des mots dans l'épigraphie médiévale n'a souvent pas de correspondance avec les unités graphiques : comme nous l'avons vu dans le chapitre réservé aux espaces, pendant les premiers siècles du Moyen Âge souvent les mots n'étaient pas précédés ou suivis d'un espace de manière régulière.

L'iconographie pouvait s'interposer aussi dans le déroulement de l'écriture, car la première était toujours réalisée avant que cette dernière, et si l'espace réservé aux lettres n'était pas suffisamment grand, l'auteur devait adopter des solutions nouvelles (comme dans la figure 29, dans laquelle le « A » de « TROECIA » ne se trouve pas à côté du « I », mais en bas, car la figure du saint ne laissait pas assez de place pour tout le mot).

Figure 29 - HS II, 75. *Cliché CIFM / CESCM*

Figure 30 - *CIFM*, HS II, 57, détail. *Cliché CIFM / CESCM*

D'autre part, il pouvait y avoir aussi des autres empêchements physiques externes, tels que les irrégularités dans la surface d'écriture (comme dans la figure 30, où le mot « directis » est coupé entre le « c » et le « t » par des irrégularités de la pierre).

La solution pour conserver à la fois cette séparation physique et la séparation grammaticale, ce n'est pas d'individualiser les mots en se servant des espaces et de la ponctuation comme dans notre orthographe moderne, mais de bien distinguer l'encodage des espaces physiques existants, l'encodage de la ponctuation et l'encodage des mots à proprement parler (après, pour la visualisation de la notice, si les mots sont bien encodés, on pourra générer automatiquement les espaces entre les mots).

Ainsi, le mieux sera de les encoder individuellement avec la balise qui leur est réservée dans la TEI : `<w>` (« word »).

Ainsi, par exemple, l'encodage du mot « DIRECTIS » de la figure 30 serait le suivant :

```
<w>DIREC<space/>TIS</w>
```

Dans les cas où il y a un saut de ligne qui coupe le mot, il s'agira tout simplement d'utiliser la balise `<lb>` avec l'attribut `@break='no'`, comme dans la figure 29 :

```
<w>TROECI<lb n=2 break='no'/>A</w>
```

Ce modèle d'encodage aide à résoudre aussi les cas où, même dans l'orthographe moderne, les mots ne sont pas définis par des espaces. C'est ce qui se passe avec les périphrases verbales et les mots enclitiques :

- Les périphrases verbales sont des mots divisés en deux parties, qui sont séparées par un espace. Pour les encoder, il ne faudra faire rien de spécial, juste marquer les limites de la périphrase avec la balise <w>. Le traitement sera le même que pour n'importe quel autre mot. En voici un exemple : `<w>vocatus sum</w>`
- Les mots enclitiques sont des mots qui sont collés à la fin du mot antérieur. Dans les

inscriptions en latin, il est assez fréquent de trouver le mot enclytique « -que ». Il sera à encoder également avec la balise <w>, comme le reste de mots, en le distinguant du mot auquel il est collé[16] : `<w>dedit</w><w>que</w>`

Lemmatisation

Par ailleurs, comme nous l'avons déjà exposé dans le chapitre consacré à l'indexation, l'encodage du mot comportera également sa lemmatisation.

La TEI et EpiDoc prévoient deux attributs différents réservés à la lemmatisation des mots : `@lemma` et `@lemmaRef`. Le premier est pensé pour accueillir directement le lemme dans son contenu, et le deuxième pour pointer vers un lien externe.

`@lemma` est le plus simple. Cependant, cette extrême simplicité ne permet pas d'enrichir l'encodage avec des informations complémentaires outres que le lemme (comme la langue du mot ou la catégorie grammaticale), ce qui, par ailleurs, peut poser des problèmes d'ambiguïté (car elle ne permet pas de distinguer les mots homonymes).

`@lemmaRef` reste beaucoup plus ouvert, et devrait permettre beaucoup plus de possibilités (telles que renvoyer à la même référence pour tous les mots qui ont le même lemme ; ou inclure des informations supplémentaires). Il peut être utilisé de deux manières :

- En pointant vers un dictionnaire externe en ligne. Pour le latin, le mieux serait d'utiliser le *Glossarium mediæ et infimæ latinitatis* de Du Cange (dans la version éléctronique mise en place par l'École des chartes). Le principal problème de cette option est sans doute l'interopérabilité des formats, car les sites web ont été pensés pour la consultation, mais pas pour la réutilisation des données (du moins directement), et à cause de cela extraire des informations concrètes systématiquement de chaque lemme paraît une mission extrêmement compliquée.
- La deuxième option serait de créer manuellement une liste de lemmes au fur et à mesure qu'on avance dans l'encodage des notices. Cela aurait l'avantage de s'assurer d'avoir toutes les informations nécessaires pour le corpus, car il serait fait sur mesure. Par contre, il comporterait un travail d'encodage manuel (même si au fur et à mesure, les noms devraient se répéter et il y

16 Ensuite, pour le traitement automatique qui permettra de générer une version normalisée du texte, il faudra avoir l'information qu'il s'agit effectivement d'un mot enclitique, qui doit être collé au mot antérieur dans l'orthographe moderne. Nous verrons comment indiquer cette information dans la partie qui suit, consacrée à la lemmatisation.

aurait beaucoup moins de travail à faire).

Cette dernière option nous a paru la plus adéquate, car elle était beaucoup plus flexible par rapport à nos besoins, et surtout elle était faisable techniquement. D'autre part, rien n'empêche d'extraire les informations de la lemmatisation des dictionnaires choisis comme référence (le Du Cange pour le latin et le *Lexique de l'ancien français* de Frédéric Godefroy (1901) pour le français).

Mais, où insérer cette liste de lemmes ? Nous avons considéré deux options possibles :

- Dans le même document, en considérant les lemmes comme une liste de mots-clés (dans `<textClass>` → `<keywords>` → `<list>`).
- Dans un document XML externe.

Il nous a semblé que le plus adéquat (et aussi le plus pratique) était de le faire dans le même document, car le principal intérêt pour que la liste soit externe était de la faire commune à toutes les notices. Mais, en fait, il existe la possibilité qu'elle soit au même temps interne et commune : en la transformant en une entité.

Les informations principales à inclure pour chaque mot lemmatisé seraient les suivantes :

- Naturellement, le **lemme**. Pour les mots avec flexion, il suivrait les mêmes règles que pour les index du *CIFM* :

Cat.grammaticale	**LATIN**	**FRANÇAIS**
Noms	En nominatif singulier	En singulier
Noms propres	En nominatif. Singulier ou pluriel en fonction du nom propre.	En nominatif. Singulier ou pluriel en fonction du nom propre.
Adjectifs	En nominatif masculin singulier	En masculin singulier
Verbes	En infinitif	En infinitif

- La **catégorie grammaticale**.
- La **langue** du mot.
- Un **identifiant**, constitué par l'initiale de la langue (« l » pour latin et « f » pour français), un tiret et le lemme.

À partir de toutes ces informations, nous avons élaboré un modèle d'encodage des lemmes comme celui de l'exemple :

```
<item xml:id='l-agnus' xml:lang='lat'>
        <w type='nom'>agnus</w>
</item>
```

Et, pour pointer vers le lemme dans le texte, il suffira d'utiliser l'identifiant dans `@lemmaRef` :

```
<w lemmaRef='#l-agnus'>agnus</w>
```

Dans le cas où il y aurait plusieurs lemmes homonymes, les informations sur la langue et la catégorie grammaticale devraient permettre de faire une première désambiguïsation. Si la langue et la catégorie grammaticale sont aussi les mêmes, nous avons pensé à une deuxième solution, qui consiste à 1) ajouter un numéro à la fin de chacun des identifiants ; 2) introduire une balise `<desc>` dans `<item>` avec une petite description du signifié de chacun des mots :

```
<item xml:id='l-esse-1' xml:lang='lat'>
            <w type='verb'>esse</w>
            <desc>Être.</w>
</item>
<item xml:id='l-esse-2' xml:lang='lat'>
            <w type='verb'>esse</w>
            <desc>Manger.</w>
</item>
```

D'autre part, nous avons considéré également pertinent de distinguer aussi les noms propres avec une balise `@subtype` ayant comme contenu `'propre'`. Cela devrait permettre de distinguer et indexer les noms propres, ce qui peut être très utile, par exemple, pour l'édition normalisée, où ils devront commencer par majuscule.

Enfin, on signalera également les mots enclitiques avec l'attribut et la valeur `@subtype='enclitique'`. Pour l'édition normalisée, cela devrait nous permettre d'éviter l'espace entre ce mot et celui qui le précède.

Forme du mot

Comme déjà avancé dans le chapitre sur l'indexation, pour compléter les informations données par le lemme, et afin d'augmenter les possibilités de la recherche en plein texte, nous avons décidé d'introduire une indexation des mots tels qu'ils se trouvent dans le texte, en conservant la flexion

grammaticale et la graphie, mais en les libérant des contraintes formelles et physiques :

- Suppression des allographes (on ne conservera que les lettres, et on les transcrira en minuscules) ;
- Développement des abréviations (et suppression des signes d'abréviation) ;
- Reconstitution des lettres perdues (si possible), sans aucun type de balisage.

Cette fonction n'est pas prévue par les attributs de la TEI, et nous n'avons trouvé aucun attribut qui lui convenait.

Finalement, grâce à l'aide précieuse de l'un des pères fondateurs de la TEI, Lou Burnard, avec lequel nous avons eu la chance de discuter de cette question, nous avons trouvé la solution qui s'adapte le mieux à nos besoins : il s'agit d'insérer, dans chaque balise `<w>`, une balise `<choice>` avec deux enfants : une balise `<orig>`, contenant toutes les informations nécessaires concernant la description du mot, et une balise `<reg>`, qui contiendra uniquement la forme du mot telle que nous venons de la décrire. Il est vrai que cette solution est assez lourde, mais c'est la meilleure que nous avons trouvée, et elle répond parfaitement à nos besoins tout en respectant les règles d'encodage de la TEI.

Exemple d'encodage d'un mot entier

Pour en terminer avec les mots, il peut être utile de résumer tout le travail d'encodage que nous venons de décrire avec un exemple pratique : celui de la figure 31, le mot « REGE » (l'ablatif singulier de « rex » en latin).

Figure 31 - *CIFM,* HS II, 2, l.2. *Cliché CESCM/CIFM*

Les informations à encoder sont les suivantes :

- « RE » est parfaitement lisible, « G » est lisible mais endommagé, et le « E » final n'est pas lisible, mais il peut être reconstruit par hypothèse à partir du contexte :

  `<orig>RE<damage>G</damage><supplied>E</supplied></orig>`
- La forme du mot est « rege » : `<reg>rege</reg>`
- Le lemme est « rex », qui est un nom latin : `<w lemmaRef='#l-rex'>` dans le texte, et la référence dans la liste de lemmes, `<item xml:id='l-rex' xml:lang='lat'><w type='nom'> rex</w></item>`.

À partir de tout cela, le résultat final sera le suivant :

```
<w lemmaRef='#l-rex'>
 <choice>
      <orig>re<damage>g</damage><supplied>e</supplied></orig>
      <reg> rege</reg>
 </choice>
</w>
```

Phrases, paragraphes et vers

Phrases

L'encodage des phrases est intéressant non seulement pour pouvoir les individualiser, mais aussi, ce qui va être indispensable dans l'édition normalisée, pour pouvoir indiquer les lettres majuscules en début de phrase (en fait, c'est la manière la plus TEI-conforme que nous avons trouvé de décrire ce type de lettres, car, vu qu'elles deviennent majuscules dans l'édition normalisée à cause du contexte syntaxique, il s'agit d'une information qui ne peut pas être encodée dans la lettre ni dans le mot, parce qu'elle provient du contexte externe).

Si les phrases apparaissent coupées entre plusieurs zones de texte (`<ab>`), entre plusieurs vers, etc., la solution sera de diviser son encodage en plusieurs balises et d'utiliser un identifiant (grâce aux attributs `@xml:id` et `@corresp`) pour les unifier.

Vers

Les vers devront être encodés dans la balise qui leur est réservée dans la TEI, `<l>`, et ils pourront être éventuellement regroupés par la balise `<lg>`, destinée à accueillir un groupe de vers.

Il faut rappeler que l'encodage des vers n'a rien à voir avec la mise en place du texte dans la surface d'écriture, même si dans certains cas la distribution par lignes peut coïncider avec les vers. Il s'agit d'une information qui appartient au style et à la composition du texte, et qui est complètement indépendante à sa matérialité.

PERSONNAGES ET LIEUX

Pour l'encodage des personnages et des lieux, la première chose qu'il faut tenir présente est la différence entre un nom propre et la référence à une personne ou lieu. Ainsi, pour la référence, la manière de citer le personnage ou le lieu peut changer même s'il s'agit toujours du même lieu ou personnage, mais le nom propre est toujours le même nom. Par exemple, « Deus », « Dominus » ou « Creator » sont trois noms propres différents qui peuvent faire référence au même personnage, Dieu. Par ailleurs, parfois la même désignation peut servir pour se référer à des personnages ou lieux différents, et ce sera le contexte qui va nous aider à le savoir.

D'autre part, le nom propre est toujours un seul mot, tandis que la référence peut être composée par plusieurs mots. Pour en citer un exemple, dans la ligne 7 de l'inscription 128 du volume HS II du *CIFM*, il apparaît la référence à Philippe VI dans l'expression « Ph(ilippe)s roy de France », une expression composée par quatre mots (dont seulement un est un nom propre de personne) qui conforment, tous ensemble, la dénomination du personnage. Il paraît donc assez évident que dans les références on n'a pas affaire à des mots concrets et figés, mais à des idées.

Ainsi, de la même manière que pour les différents allographes d'une même lettre il nous a fallu faire le lien manuellement pour nous assurer que chaque allographe pointait toujours vers le caractère correct, ici il s'agira aussi de faire une liste de référence des lieux et des personnages pour ensuite pouvoir les identifier lorsqu'on les retrouvera dans le texte. La place réservée à cela dans la notice est, comme nous l'avons déjà signalé plus haut, dans `<textClass>`, et chacune des listes sera encodée dans une entité.

Le nom donné à chaque personnage et à chaque lieu sera le nom régularisé en français moderne (en suivant le modèle suivi par les index des éditions papier du *CIFM*), et la balise devra contenir un identifiant.

En voici un exemple :

```
<textClass>
    <keywords>
        <list type='personnages'>
            <item xml:id='Simon-apôtre'>
                <forename>Simon</forename>, apôtre
            </item>
            (...)
```

```
                </list>
            </keywords>
        <textClass>
```

Enfin, pour les citer dans le texte, il faudra garder la distinction entre personnages et lieux. Ainsi, les références aux personnages devront être balisées avec `<persName>`, tandis que les références aux lieux, avec `<placeName>`. Dans tous les cas, il sera indispensable l'attribut `@corresp`, qui pointe vers le nom de la référence dans la liste.

Pour continuer avec l'exemple antérieur, voici comment la référence à l'apôtre Simon serait encodée dans le texte :

```
    […]  IN  HONORE  SANCTORVM  APOSTOLORVM  <persName  corresp='#Simon-apôtre'>SYMONIS</persName> ET IVDAE ET OMNIVM APOSTOLORVM […]
```

Niveau 4 : <div @type='translation'> - *Traduction du texte épigraphique*

La deuxième partie du `<body>` est destinée à contenir la traduction du texte épigraphique, en suivant les normes d'édition que nous avons définie.

Le travail d'encodage sera extrêmement simple, et il suffira d'indiquer les paragraphes (`<p>`) ou les vers (`<l>`) en fonction du type de texte.

Niveau 4 : <div @type='comment'> - *Commentaire sur l'inscription*

Cette balise devra contenir le commentaire sur l'inscription. Le balisage sera également très simple, et il se limitera seulement à la division par paragraphes (`<p>`), à l'encodage des notes de bas de page (`<note place='bottom'>`) et au signalement des citations (`<q xml:lang='...'>` pour les mots et constructions syntaxiques, et `<c>` pour les caractères).

Niveau 4 : <div @type='bibilography'>

Enfin, la dernière partie du `<body>` est réservée à la bibliographie. Elle contiendra la bibliographie sur l'inscription, mais elle peut accueillir aussi d'autres types de bibliographie. Pour l'instant, nous avons pensé à l'utiliser aussi pour faire des liens avec la base de données d'iconographie romane du CESCM, ROMANE, dans les cas où l'iconographie qui accompagne les inscriptions est décrite dans cette base de données.

Pour distinguer ces différents types de bibliographie, il s'agira de définir une liste bibliographique différente (`<listBibl>`) pour chaque cas, et de lui donner un nom à travers de l'attribut `@type`. Ainsi, la bibliographie sur l'inscription extraite de la notice sera à encoder dans `<listBibl type="inscription">`, tandis que la bibliographie de la base Romane, dans `<listBibl type="Romane">`.

Pour l'encodage des références bibliographiques (`<bibl>`) nous nous sommes fondés essentiellement sur les recommandations de la TEI :

- L'auteur sera encodé dans la balise `<author>`.
- Le titre, dans la balise `<title>`. Ici, nous avons considéré pertinent d'introduire un attribut `@type` qui peut contenir les valeurs `'book'` ou `'article'` en fonction du type d'item bibliographique.
- La date, dans la balise `<date>`.
- La page, le chapitre ou le volume dans la balise `<biblScope>` et l'attribut `@unit` accompagné de la valeur correspondante dans chaque cas (`'page'` pour page, `'chapter'` pour chapitre et `'volume'` pour volume).
- Les images, dans `<figure>` et ensuite `<desc>`.
- Les URIs, comme valeur de l'attribut `@target` dans la balise `<link>`.

Voici l'exemple d'encodage de la bibliographie dans la notice 89 du volume HS II du *CIFM*, qui contient aussi des références vers la base Romane :

```
<div type="bibliography">
    <listBibl type="inscription">
        <bibl>
            <author>Crozet</author>
            <title type="article">La crypte de Notre-Dame-la-
            Grande</title>
            <date when="1940">1940</date>
            <biblScope unit="page">435 (texte)</biblScope>
        </bibl>
        <bibl>
            <title type="book">CIFM</title>
            <biblScope unit="volume">1</biblScope>
            <date when="1974">1974</date>
            <biblScope unit="page">17-18</biblScope>
```

```
                <biblScope unit="chapter">18</biblScope>
                <figure><desc>fig. 20-22</desc></figure>
            </bibl>
            <bibl>
                <author>Hulnet-Dupuy</author>
                <title type="article">Les peintures
                murales</title>
                <date when="2002">2002</date>
                <biblScope unit="page">219-220</biblScope>
                <figure><desc>fig. 278-280</desc></figure>
            </bibl>
        </listBibl>

        <listBibl type="Romane">
            <bibl>
                <title>Voûte de la crypte</title>
                <biblScope unit="chapter" n="3422">3422
                </biblScope>
                <link target="http://romane.critt-
                informatique.fr/visualise.aspx?fiche=3422"/>
            </bibl>
            <bibl>
                <title>Voûte de la crypte, partie centrale</title>
                <biblScope unit="chapter" n="3423">3423
                </biblScope>
                <link target="http://romane.critt-
                informatique.fr/visualise.aspx?fiche=3423"/>
            </bibl>
        </listBibl>
    </div>
```