



HAL
open science

Manuel d'encodage XML-TEI - édition numérique de manuscrits baroques

Nicole Dufournaud, Valérie Gratsac Legendre

► **To cite this version:**

Nicole Dufournaud, Valérie Gratsac Legendre. Manuel d'encodage XML-TEI - édition numérique de manuscrits baroques : Recommandations pour une application TEI. 2012. hal-00718043

HAL Id: hal-00718043

<https://hal.science/hal-00718043>

Submitted on 16 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Manuel d'encodage XML-TEI – édition numérique de manuscrits baroques

EHESS - CRH



Recommandations pour une application TEI – Projet ENBaCH

Nicole Dufournaud et Valérie Gratsac-Legendre

Version du 12 juillet 2012

Table des matières

| | |
|--|----|
| Remerciements..... | 4 |
| En préambule..... | 5 |
| Introduction..... | 6 |
| Les principes de balisage..... | 6 |
| La création de fichier..... | 6 |
| 1Fichier type..... | 7 |
| 2Nom de fichier..... | 7 |
| Les métadonnées..... | 7 |
| 1Définition..... | 7 |
| 2L'en-tête de la TEI..... | 8 |
| Les pointeurs, les liens, les ancres, les segments, les références croisées..... | 8 |
| Le processus d'encodage..... | 9 |
| 1Les éléments retenus..... | 10 |
| 2Les attributs retenus..... | 11 |
| 3La structure minimale..... | 12 |
| 4La prose..... | 13 |
| 5Les vers..... | 14 |
| 6Les listes..... | 14 |
| 7Destruction et restitution..... | 15 |
| 8Les blancs..... | 15 |
| 9Les annotations..... | 15 |
| Les notes en marge de gauche du projet Marillac..... | 16 |
| Les notes en marge de gauche du projet Patin..... | 16 |
| 10Les suppressions et les ajouts..... | 16 |
| Ajouts..... | 17 |
| Ratures..... | 20 |
| Le rendu..... | 22 |
| Suppressions et ajouts concomitants | 22 |
| Avec <delSpan> et <addSpan>..... | 24 |
| 11Numérotation multiple des pages et des feuillets..... | 26 |
| Le balisage sémantique..... | 26 |
| 1Les langues..... | 26 |
| 2Les références bibliographiques..... | 27 |
| 3Discours et citations..... | 28 |
| Les discours oraux..... | 28 |
| Les citations écrites..... | 30 |
| 4Les entités nommés..... | 34 |
| La temporalité..... | 34 |
| Les noms de personnes et de lieux..... | 37 |
| 5L'indexation..... | 39 |
| 6Les références croisés..... | 40 |
| Index et renvois..... | 40 |
| Liens et références..... | 41 |
| La correspondance et l'alignement entre les textes..... | 42 |
| Identifier chaque paragraphe..... | 42 |
| Les liens..... | 43 |
| L'analyse interprétative des textes..... | 44 |
| Conclusion..... | 44 |

| | |
|--|----|
| Annexe 1 : fichier type du projet Marillac..... | 45 |
| Annexe 2 : la feuille de style de travail (extrait)..... | 50 |
| Annexe 3 : la grille d'analyse du projet Marillac (extrait)..... | 55 |
| Annexe 4 : liste des « valeurs » d'attributs non variables..... | 57 |
| Liste des éléments et attributs..... | 59 |

Remerciements

L'édition d'un manuel est le fruit d'un long travail de réflexion qui doit aussi beaucoup à trois personnes que nous tenons à remercier ici.

- Les citations et les discours ont posé beaucoup de problèmes d'encodage. Dora de Lima (Université Paris I – Panthéon Sorbonne, CRHM) a accepté de collaborer et d'assister à des séances de travail. Nous lui devons une partie des résultats.
- Lou Burnard (MEET, TGE Adonis) a précisé quelques encodages que nous avons définis préalablement.
- Enfin, Jean-Daniel Fekete, directeur de recherche à l'INRIA, responsable de l'équipe-projet AVIZ, nous a apporté une aide précieuse tout au long de la mise en place du projet d'édition et de l'élaboration du manuel.

Qu'ils en soient tous les trois remerciés ici !

En préambule

Le projet européen « ENBaCH »¹ comprend plusieurs projets dont deux sont réalisés en XML/TEI ; ce sont les projets d'édition numérique des manuscrits suivants :

1. *La Vie de Michel de Marillac* par Nicolas Lefebvre de Lezeau
2. *Les Papiers* de Guy Patin – manuscrit de Vienne

Ce manuel d'utilisation de la TEI a été rédigé au fur et à mesure de l'avancement des projets et sur la base des Guidelines de la TEI en ligne sur <http://www.tei-c.org/>. Il est complété par les annexes et un index des éléments et attributs utilisés.

L'encodage est à dissocier du rendu formel en HTML qui est du domaine des feuilles de style en XSLT et de la CSS que nous ne développerons pas ici ; seul un extrait de la feuille de style de travail en XSLT est annexée en fin de manuel.

Pour la compréhension du concept de balisage et de l'encodage, nous recommandons le manuel d'utilisation des Bibliothèques Virtuelles des Humanistes sur <http://www.bvh.univ-tours.fr/index.asp>.

Établi entre septembre 2010 et juin 2012, ce document ne doit pas être tenu comme le résultat de l'encodage du projet ENBaCH : il représente l'état d'avancement d'un projet à mi-parcours.

Enfin, ce manuel est avant tout le fruit de la collaboration entre les deux autrices.

¹ Sur le projet ENBaCH, voir les sites <http://www.enbach.eu/EN.aspx?DefaultLanguage=EN> et <http://enbach.hypotheses.org/>

Introduction

Les deux sous-projets du projet ENBaCH suivent les recommandations du consortium TEI (<http://www.tei-c.org/>) dans sa version P5 qui permettent le respect de la présentation originale : pagination, lignage, disposition des titres, des paragraphes et des notes. Ce respect quasi-diplomatique des documents autorise des traitements informatiques ultérieurs que ce soient sur les langues, les citations, l'alignement des paragraphes, les analyses historiques et les variantes du texte.

Brièvement rappelons qu'un texte a une structure physique, logique et sémantique qui influe sur l'encodage. En conséquence, nous différencions trois types de balisage :

1. le balisage physique, formel et typographique encode les structures logique et formelle du texte ;
2. le balisage sémantique autorise un encodage d'un premier niveau d'analyse comme les noms, les toponymes, les citations, les discours et les dates ;
3. le balisage analytique complète l'apparat savant de l'encodage.

Les principes de balisage

La TEI est une norme de codage de textes qui repose sur XML². Sans rappeler les notions de base de ce meta-langage, nous insistons sur deux particularités :

- l'imbrication d'éléments dans d'autres éléments, les premiers étant les « enfants » des seconds, dits « parents » ;
- l'interdiction de faire chevaucher les éléments.

Les éléments sont de deux types :

1. ceux qui définissent la structure du document et du texte ;
2. ceux qui permettent une analyse soit succincte comme le balisage sémantique soit plus approfondie comme le balisage analytique.

Associés aux éléments, les attributs sont de deux types :

1. l'attribut global ;
2. l'attribut spécifique à un élément.

A l'attribut, est associée une valeur.

Les éléments et des attributs avec leurs valeurs sont indiqués sous la forme suivante :

```
élément attribut="valeur" attribut="valeur"
```

La création de fichier

La création de fichier dépend du logiciel d'édition choisi. Pour notre confort personnel, nous éditons sous deux logiciels : Emacs et Eclipse. Pour ce dernier, le plugin « Millefeuille » a été retenu : il a été développée sous forme d'un plugin à [Eclipse](#).

Emacs est un éditeur de texte très populaire et très puissant parmi les programmeurs en informatique.

² XML pour *Extensible Markup Language* est un langage informatique dont l'ancêtre est SGML. Sa syntaxe autorise son extension grâce à une grammaire définie dans un schéma.

Eclipse est une plate-forme de développement et d'édition bien connu dans le milieu du développement informatique. Elle possède de nombreuses fonctionnalités pour le travail sur des documents [XML](#) de façon coopérative. Le plugin développé ajoute des fonctionnalités d'aide à l'édition, d'affichage et permet l'enrichissement en information des documents encodés en XML ainsi que la navigation dans les fichiers grâce à un système d'index. Le plugin est entièrement configurable et extensible à d'autres projets d'encodage au format XML³.

1 Fichier type

Selon l'éditeur, la création de fichier peut se faire automatiquement et ne pose pas de problème particulier. Dans Millefeuille, un « fichier type » ou « *defaultContents* » est obligatoire. Ce fichier peut être modifié (voir annexe 1).

2 Nom de fichier

Les noms de fichiers doivent être construits. Il n'existe pas de règle ; nous donnons pour exemple celui du projet Marillac ci-après.

A partir des chapitres existants dans les trois manuscrits des Archives nationales, de la bibliothèque Sainte Geneviève et de la bibliothèque nationale de France, nous avons élaboré une liste ordonnée de fichiers : un fichier correspond soit à un chapitre, soit aux pièces liminaires, soit aux annexes. Chaque fichier porte le lieu de conservation, le numéro ordonné du chapitre et le nom abrégé du chapitre ce qui donne la forme suivante :

Pour les Archives nationales : AN_chap_numero_nom.xml

Pour la Bibliothèque Sainte Geneviève : BSG_chap_numero_nom.xml

Pour la Bibliothèque Nationale de France : BNF_chap_numero_nom.xml

Les métadonnées⁴

Préalablement à tout encodage, il est nécessaire de donner des informations descriptives et déclaratives sur le texte en vue de sa diffusion, d'un échange de données ou encore de son exploitation : ces composantes – ou métadonnées – forment l'en-tête TEI nommé `<teiHeader>`. Réfléchir à la structure d'un document permet également de renseigner cet entête.

1 Définition

Les métadonnées peuvent être définies comme des données servant à représenter ou à décrire d'autres données, ici les deux textes édités. Elles contiennent des informations sur la source d'un document, sa nature, son contenu, sa localisation, son histoire, son statut juridique, etc. Elles remplissent un rôle d'indexation qui facilite l'accès au contenu et améliore la recherche. Elles doivent être normalisées grâce à des standards comme Dublin Core, EAD (Encoded Archival Description) par exemple afin de permettre l'interopérabilité et l'échange des données.

La métadonnée est le nom de la notice électronique de contenu du texte comme l'était la fiche cartonnée dans les bibliothèques ou les archives, il y a peu de temps encore.

³ Voir le manuel de Félicien François sur <http://millefeuille.gforge.inria.fr/>

⁴ Cette section sur les métadonnées reprennent en grande partie le cours donné par Laetitia Bontemps lors de la formation XML/TEI au Centre d'Études Supérieures de la Renaissance à l'Université de Tours, octobre 2011.

2 L'en-tête de la TEI

La TEI est un format d'encodage standard qui comprend des métadonnées situées dans la partie qui s'intitule `<teiHeader>`.

Dans la structure globale d'un fichier XML/TEI, ce qui est appelé communément le « header » est dissocié du texte. Le « header » que nous nommerons dorénavant l'en-tête TEI présente une description structurée des données contenues dans le fichier XML/TEI. La hiérarchie des éléments TEI est contrainte par le schéma choisi.

L'en-tête TEI – en état actuel des projets – sera développé dans la seconde phase du projet ENBaCH, entre 2012 et 2014.

Les pointeurs, les liens, les ancres, les segments, les références croisées

Les deux projets nécessitent d'utiliser des liens, des références croisées et des pointeurs qui obligent à identifier des segments de textes pour les relier entre eux. Nous abordons ici de façon succincte ces notions qui seront ensuite développées dans chaque partie.

Nous dressons la liste exhaustive des éléments et attributs de ce groupe :

Eléments

- pointeur simple : `<ptr>`
- références croisées : `<ref>`
- liens : `<join>`, `<link>` et `<linkGrp>`
- segments : `<seg>`
- ancres : `<anchor>`

Attributs possibles :

- pointeur : `@target`, `@type`, `@who` et `@spanTo`
- identifiant : `@xml:id`

Seul ce dernier attribut identifie la cible du pointeur.

Parmi les nombreux pointeurs, notre intérêt se porte dans un premier temps vers ceux qui pointent vers un autre emplacement comme l'élément `<ptr>` et ceux qui délimitent une zone du texte avec un début et une fin comme `<ref>`. Parmi les attributs, `@target` en précise la destination et `@type`, la typologie.

Dans le projet Marillac, ces pointeurs sont en premier lieu utilisés pour marquer les ratures et les ajouts, le point d'ancrage `<anchor>` délimitant la fin de la zone concernée. Les attributs utilisés sont alors `@spanTo` et `@xml:id`⁵. Dans un second temps, nous les utilisons dans le balisage sémantique et principalement dans l'encodage des discours⁶.

Pour le projet Patin, l'utilisation des références croisées prime sur les autres éléments.

⁵ Pour le développement de ces pointeurs, voir la partie du manuel sur les ratures et les suppressions.

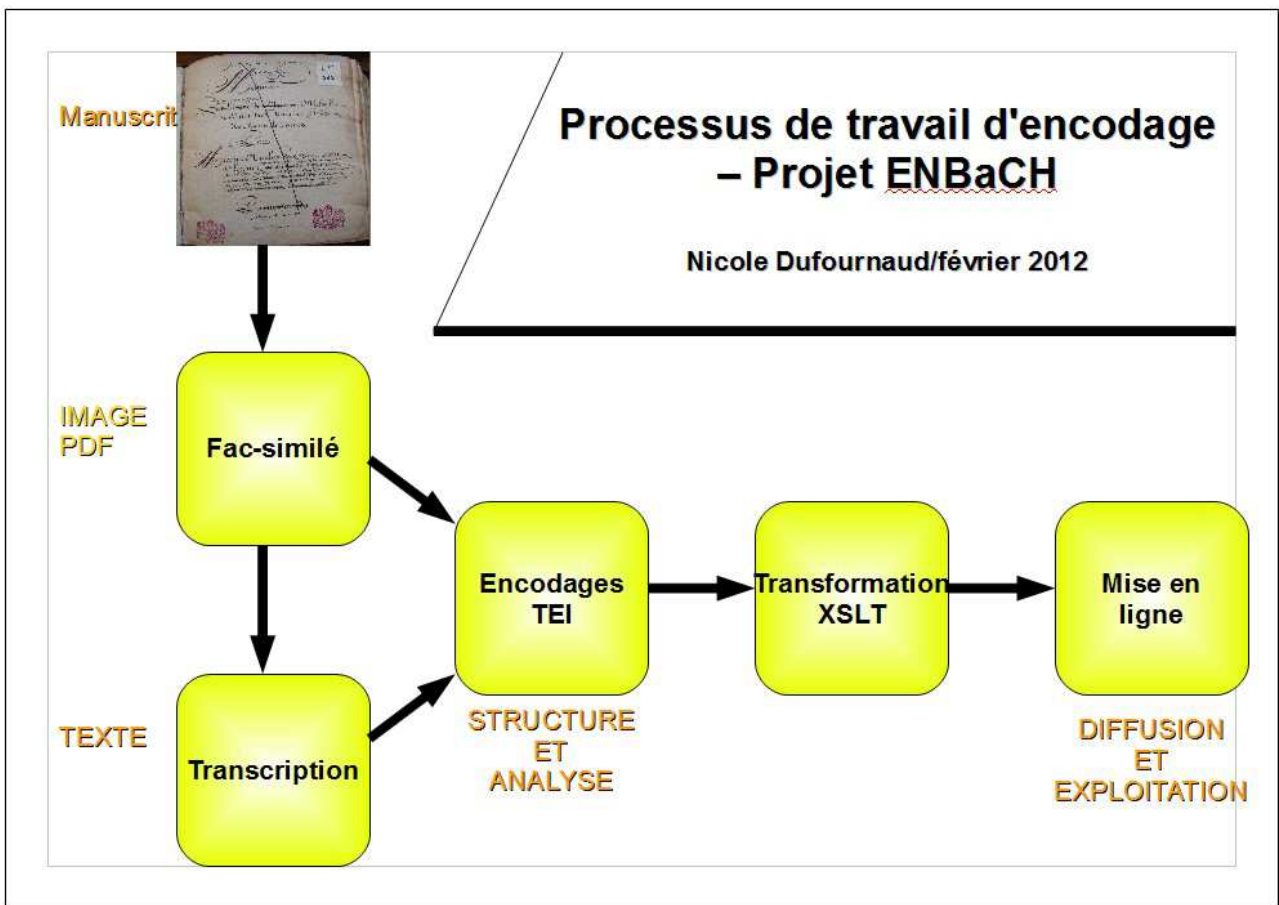
⁶ Pour le développement de ces pointeurs et liens, voir la partie du manuel sur les discours.

Le processus d'encodage

Le processus du travail d'encodage permet d'obtenir un texte structuré dont les données pourront être traitées par un ordinateur. Afin que le résultat soit cohérent, la rigueur impose de suivre la structure logique et physique du document, puis la mise en forme du texte ; à partir de cette première édition, il s'agit d'opérer un choix sur les index à produire afin d'encoder les éléments sémantiques comme les nom de lieux, de personnages, les dates, etc.

Prenons l'exemple de nos deux projets. Après une première numérisation du manuscrit récupérée sous forme de fac-similés en format PDF, le premier travail est de transcrire les textes. Il s'agit d'attendre patiemment la fin des contrôles de la transcription avant de commencer le travail d'encodage, cette phase de pré-encodage étant capital pour la bonne qualité de l'édition finale. Les textes étant récupérés, il s'agit alors de les structurer grâce à l'encodage XML/TEI. Dans une première passe, la structure logique et physique du document avec la mise en forme est encodée ; une attention particulière permet l'encodage des nombreuses ratures et ajouts, ainsi que des blocs de ratures avec leur signification. Dans un seconde passage, une lecture attentive du texte permet d'encoder les mots analysés comme les noms de personnages, les dates mais aussi les citations et les discours ; cet encodage permet la production d'index. L'édition numérique est alors prête pour être mise en ligne⁷.

Le schéma suivant montre les différentes étapes dans le processus du travail d'encodage des deux projets.



⁷ Pour introduire ces notions, il est recommandé de consulter les articles de Jean-Daniel Fekete et Nicole Dufournaud et particulièrement « Analyse historique de sources manuscrites : application de TEI à un corpus de lettres de rémission du XVI^e siècle », *Document numérique* « Les documents anciens », Hermes, Paris, volume 3, n° 1-2, juin 1999, en ligne sur http://www.lri.fr/~fekete/ps/doc_num_fekete_duf.pdf

Dans la partie du processus réservé à l'encodage, il s'agit dans un premier temps d'examiner la structure du document, puis dans un second temps à choisir la « grammaire » appelée « modèle » qui conviendra à l'encodage du document. Deux solutions s'offrent à nous : choisir un modèle pré-existant ou en construire un nouveau. Vu l'énorme travail effectué par les spécialistes de la TEI, nous préconisons de choisir un modèle qu'il faut compléter. Dans une dernière phase, il nous faut choisir les éléments et les attributs du modèle.

1 Les éléments retenus

Les éléments utilisés sont ici les suivants :

- `<text>` contient un texte
- `<front>` contient les pièces liminaires
- `<body>` contient le corps du texte
- `<div>` contient la subdivision du texte
- `<head>` contient le titre
- `<back>` contient les annexes

Ce qui donne la structure générale du fichier :

```

<teiHeader>...</teiHeader>
<text>
  <front>...</front>
  <body>
    <div>...</div>
  </body>
  <back>...</back>
</text>
```

Rappel : le manuscrit « Marillac » étant constitué de chapitres, nous avons donc choisi comme unité de base le chapitre pour ce projet ; à chaque chapitre correspond un fichier.

Puis nous pouvons commencer le balisage formel du corps du texte :

- `<pb>` pour les changements de page
- `<fw>` pour les en-têtes et pieds de pages
- `<p>` pour les paragraphes
- `<milestone>` pour un changement de structure
- `<floatingText>` pour l'insertion d'un texte dans le texte
- `<lb>` pour les changements de ligne
- `<abbr>` pour les abréviations
- `<expan>` pour le développement des abréviations
- `<unclear>` pour les approximations
- `<note>` pour les annotations

- `` pour les ratures
- `<add>` pour les ajouts
- `<sic>` pour les incohérences
- `<space>` pour les espaces laissés en blanc

A l'issue de cette première passe, la version HTML offre une présentation quasi-similaire à l'original.

Dans un second passage, nous analysons le texte avec les éléments suivants :

- `<quote>` pour les écrits rapportés
- `<cit>` pour les citations
- `<bibl>` pour les références bibliographiques
- `<author>` pour l'auteur
- `<title>` pour les titres
- `<q>` pour les discours
- `<foreign>` pour les langues étrangères
- `<persName>`, pour les noms de personnes
- `<placeName>`, pour les noms de lieux
- `<date>` pour les dates
- `<time>` pour les heures

Dès lors, nous proposons une frise chronologique dite « Timeline » encodé ainsi :

- `<listEvent>` contient la liste des événements
- `<event>` contient les données de l'événement dans lequel se trouvent imbriqués les éléments :
 - `<head>` contient le titre de l'événement
 - `<label>` contient l'étiquette qui sera affichée
 - `<desc>` contient la description de l'étiquette

2 Les attributs retenus

Les attributs sont soit de type global, c'est-à-dire utilisables par tous les éléments, soit de type spécifique, c'est-à-dire réservés à un ou plusieurs éléments.

Les attributs sélectionnés sont les suivants :

- `@type` pour établir une typologie
- `@place` pour indiquer un emplacement
- `@target` pour cibler
- `@unit` pour indiquer une unité

- @quantity pour indiquer une quantité

Par défaut, à chaque attribut est associée une valeur en langue anglaise recommandée par les Guidelines de la TEI. En revanche, le cas échéant, nous créons une valeur en langue française (Voir en annexe la liste des valeurs retenues). Nous donnons ici un exemple de citation d'une lettre :

```
<quote type="letter">ceci est un exemple de citation de lettre</quote>
```

L'élément <quote> marque la citation alors que la valeur "letter" de l'attribut @type précise le type de la citation : ici une lettre.

3 La structure minimale

Le texte a une structure minimale composée de pièces liminaires puis d'un corps de texte et enfin de pièces annexes. Le corps du texte est composé de division (partie, chapitre, etc.), elle-même regroupant des blocs de textes comme des strophes ou des paragraphes. La TEI en restitue la structure : nous prenons ici l'exemple des paragraphes.

```
<text>
  <front>
</front>
  <body>
    <div>
      <p>blablabla</p>
    </div>
  </body>
  <back>
</back>
<text>
```

Avant le corps du texte, les pièces liminaires peuvent notamment contenir une page de garde balisée grâce aux éléments <titlePage>, <docTitle> et <titlePart> dans <front> comme dans l'exemple suivant tiré du chapitre « Mortifications » :

```
<front>
  <titlePage>
    <docTitle>
      <titlePart>
        <lb/>Mortifications
      </titlePart>
    </docTitle>
  </titlePage>
</front>
```

Parfois, un mot ou une phrase termine la division comme le mot « FIN ». L'élément <trailer> est alors nécessaire :

```

<body>
  <div>
    <p>...</p>
  </div>
  <trailer>FIN</trailer>
</body>

```

Quand un document est inséré dans le discours, la structure est plus compliquée et nécessite le recours à l'élément `<floatingText>` comme dans l'exemple suivant tiré du chapitre sur la « Loi salique ».

```

<div>
  <p>...</p>
  <floatingText>
    <body>
      <p>
        <quote type="decision">
          <lb/>Sur la Remonstrance cy devant faicte à la cour
          <lb/>par le Procureur general du Roy, et la matiere
          <lb/>mise en deliberation, ladicte cour toutes les
          <lb/>chambres assemblées, n'ayant, comme elle n'a
          <lb/>jamais eu, aultre intention, que de maintenir
          <lb/>la Religion Catholique, Apostolique, et Romaine,
          <lb/>et l'Estat, et Couronne de France, soubz la
          <lb/>protection d'un Roy tres chrestien, Catholique
          <gap/>
        </quote>
      </p>
    </body>
  </floatingText>
  <p>...</p>
</div>

```

4 La prose

Le texte en prose est structuré avec les éléments `<p>` pour le paragraphe et `<lb/>` pour la ligne. Par exemple la phrase suivante est encodée ainsi :

```

<p>
  <lb/>Et Louis de Marillac posthume, qui depuis
  <lb/>a esté mareschal de France.
</p>

```

Cependant, quand le bloc de texte n'est manifestement pas un paragraphe, il est recommandé d'utiliser l'élément `<ab>` pour *anonymous* block.

5 Les vers

Pour encoder des lignes de vers, deux autres éléments sont utilisés :

<lg> qui regroupe des lignes de vers

<l> qui balise la ligne de vers

comme dans l'exemple tiré du chapitre « De ses amis » (Marillac).

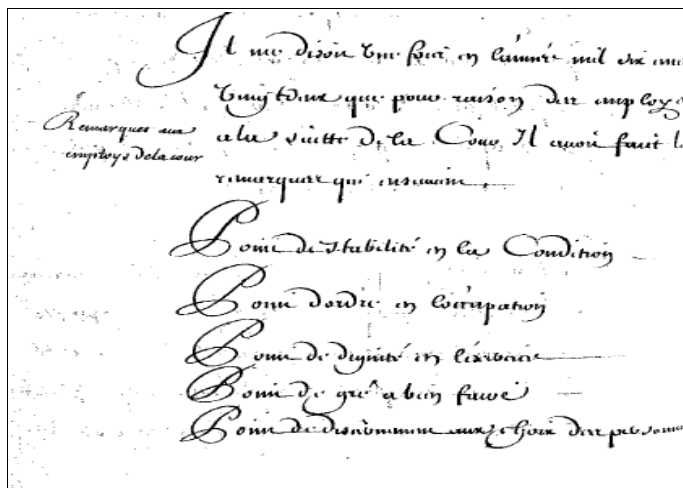
```
<lg>
<l>Faelices animae vestri duo lumina faeli secli</l>
<l>O ames bienheureuses</l>
<l>de vostre temps lumieres pretieuses</l>
</lg>
```

Dans une citation, il arrive qu'il y ait une seule ligne de vers : elle est alors insérée dans le paragraphe dans l'élément structurant la citation (voir la partie sur les citations).

6 Les listes

Parfois, la lecture d'un texte fait apparaître une structure peu visible et pourtant bien connue : la liste. Elle comprend des unités appelées « item » ordonnées ou non, et parfois une « étiquette ». Les éléments utilisés sont les suivants : <list> <label> et <item>. L'élément <head> est parfois requis. Les attributs sont les suivants : @n et @type avec une valeur « simple », c'est-à-dire non classée et non marquée.

Voici un exemple tiré du chapitre 17 du manuscrit Marillac (version BNF, page 102 verso) avec l'encodage de la liste.



Il en résulte l'encodage suivante :

```
<p>
<lb/>Il me disoit une fois en l'année mil six cens
<lb/>Ving deux que pour raison des employs
<lb/>à la suite de la Cour Il avoit faict les
<lb/>remarques qui ensuivent
```

```

</p>
<list type="simple">
  <item><lb/>Poinct de stabilité en la condition</item>
  <item><lb/>Poinct d'ordre en l'occupation</item>
  <item><lb/>Poinct de dignité en l'exercice</item>
  <item><lb/>Poinct de gré à bien faire</item>
  <item><lb/>Poinct de discernement aux choix des personnes</item>
</list>

```

7 Destruction et restitution

La structure physique intervient dans la transcription que ce soit en raison de dommage, de destruction et de restitution. Dans le projet Marillac, les manuscrits sont en bon état et ne nécessitent pas le recours à un encodage particulier. En revanche, dans le projet Patin, il est parfois utile de restituer du texte qui ne peut être lu par suite de dommages matériels.

Souvent la reliure du manuscrit peut empêcher la bonne transcription d'un mot. Le transcritteur restitue alors le mot et en prend la responsabilité. L'élément destiné à la restitution est `<supplied>`. On peut donner la raison du dommage grâce à son attribut `@reason`. Voici un exemple (Patin, page 10) :

```
toutes les fanfreluc<supplied reason="binding">hes</supplied>
```

8 Les blancs

Dans les manuscrits, fréquemment des blancs sont laissés pour une complétion ultérieure ou pour aérer le texte. L'élément `<space>` a été créé pour encoder ces blancs ; s'y ajoutent des attributs comme `@unit` pour donner l'unité – pixel, cm, pouce – et `@quantity` pour la quantité.

Prenons un exemple du projet Marillac : dans le chapitre sur les Offices (Marillac, page 47 verso⁸), un blanc est visible à la fin de la page entre la fin du paragraphe et un ajout.

```
il a faict. <space unit="cm" quantity="3"/>Il fust
```

Cet élément est également utilisé quand l'espace est réservé par le scribe à un usage ultérieur, un numéro de page, par exemple (Patin, page 29).

```
Ces deux hommes estoient <space unit="cm" quantity="3"/> et Philippe 4.
```

Le même élément répond donc à deux usages différents.

Remarque : l'espace laissé en blanc peut marquer un changement de paragraphe sans pour autant qu'il y ait un passage à la ligne.

9 Les annotations

Par annotations, nous entendons les notes en marge. Nous différencions ici les deux projets : Marillac et Patin. Les notes des manuscrits Marillac sont des annotations simples alors que celles de

⁸ La pagination des feuillets verso est inexistante dans le manuscrit. Elle a été construite pour faciliter le repérage.

Patin sont de deux types : le premier est une entrée de type encyclopédique et le second des notes avec des renvois.

Les notes en marge de gauche du projet Marillac

Les notes en marge de gauche sont transcrites et encodées avec l'élément `<note>` et l'attribut `@place` avec sa valeur `"margin-left"` mais ne sont encodées ni formellement ni sémantiquement⁹.

Voici un exemple d'encodage tiré du chapitre « Décès » (Marillac, page 1) :

```
<lb/>Le dimanche premier jour d'aoust feste de
<lb/><note place="margin-left">Le 1 Aoust 1632 tomba malade</note>
Saint Pierre aux liens, le temps estant venu, que
<lb/>Dieu vouloit deslivrer son ame des liens de son
<lb/>corps apres s'estre mis au lict en bonne santé,
```

Remarque : pour une meilleure lisibilité dans l'éditeur de textes, nous recommandons de présenter le texte ligne par ligne et d'isoler la note.

Les notes en marge de gauche du projet Patin

À la différence du projet Marillac, les notes en marge de gauche sont transcrites et encodées formellement et sémantiquement.

La note est encodée avec l'élément `<note>` et l'attribut `@place` avec sa valeur `"margin-left"`. Puis elle est typée avec l'attribut `@type` ; deux valeurs sont choisies : `"entry"` pour les entrées encyclopédiques et `"note"` pour les autres notes.

Voici deux exemples :

```
<note place="margin-left" type="entry">Fam. Strada.</note>
```

et deux lignes plus loin dans le même paragraphe :

```
<note place="margin-left" type="note">Voyez plus bas pag.70. et 75.</note>
```

Nous distinguons donc deux types de notes : l'annotation comprenant une référence avec un lien vers un autre emplacement dans le texte, de l'entrée de la notice sur le personnage « Strada ».

10 Les suppressions et les ajouts

Parmi les éléments de base, nous en retenons particulièrement deux qui permettent d'encoder les ratures et les ajouts observés principalement dans le projet Marillac : `<add>` et ``.

Les **ajouts** sont de deux types : un ajout simple qui va du mot à quelques lignes, puis l'ajout plus conséquent qui va jusqu'au paragraphe. Les ajouts peuvent être liés à des ratures et vice-et-versa. La

⁹ Voir le balisage sémantique.

règle générale du non-chevauchement entre éléments en XML est ici parfois difficile à respecter.

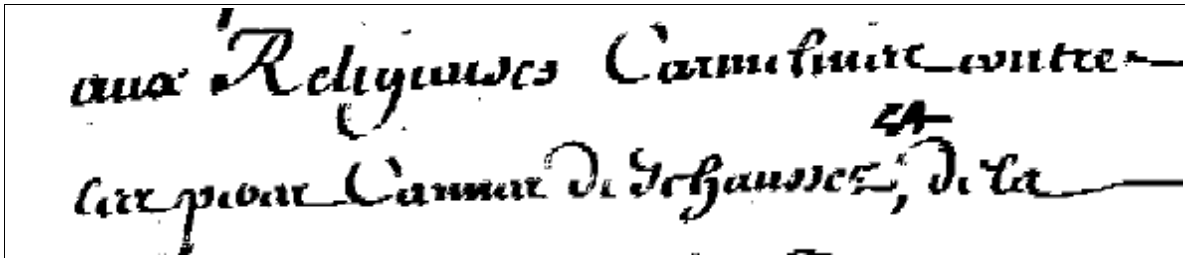
Les **ratures** sont de deux types : une rature simple qui va du mot à quelques lignes ; puis la rature plus conséquente couvrant de grandes parties de textes de la dizaine de lignes jusqu'au paragraphe voire la page. Par ailleurs, les ratures peuvent prendre plusieurs aspects formels : trait simple, circonvolution, diagonal, etc. que nous traitons grâce à un rendu raffiné.

Ajouts

L'élément <add> est utilisé pour les ajouts. Nous distinguons les ajouts sur la ligne ou au-dessus de la ligne de ceux en marge (marges de gauche, en haut et en bas). Le signe « + » présent dans le manuscrit original est respecté ; il marque l'endroit de l'ajout qui pointe dans la marge.

Pour les ajouts simples, nous avons choisi de ne pas signifier où se situe exactement l'ajout : nous privilégions la fluidité de la lecture.

- Par défaut, il se trouve soit sur la ligne soit au-dessus, comme le montre l'exemple tiré du chapitre 10 (version BN, page 230) :



que nous encodons ainsi :

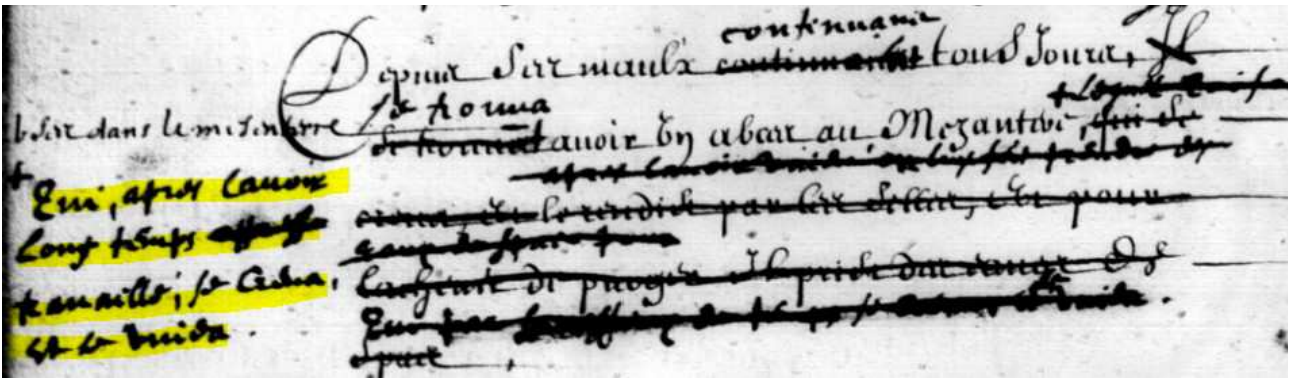
```
<lb/>aux Religieuses Carmelines contre  
<lb/>les Peres Carmes des chaussez, <add>et</add> de la
```

avec le résultat en HTML suivant :

aux Religieuses Carmelines, contre
les Peres Carmes des chaussez, et de la

Le mot « et » est en couleur et sur la ligne. Ce choix d'une lecture intuitive et linéaire peut être bien entendu remis en cause par le simple fait d'un changement dans la feuille de style.

- Pour les ajouts situés en marge, l'attribut @place avec la valeur « margin-left » est utilisé. L'exemple suivant montre un passage ajouté en marge, surligné en jaune par nos soins, (Marillac, chapitre 11, page 304 verso) :



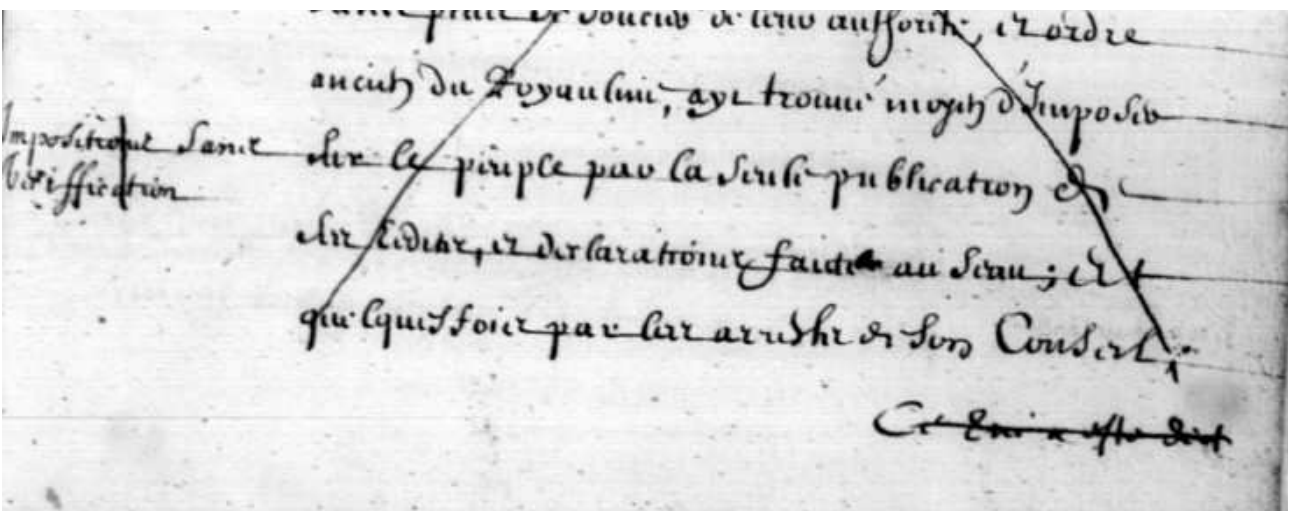
que nous encodons ainsi :

```
<add place="margin-left">qui, apres l'auoir longtemp <unclear/>
travaillé, se creuva, et se vuida.</add>
```

avec le résultat en HTML suivant :

| | |
|---|--|
| Depuis ses maux continuans toujours il se trouva auoir un abces au mezantere, qui se + ereva, et le rendist par les selles, et pour fachever de purger il prist des caues de Spas: <u>qui par</u> | Abces dans le mesentere |
| | qui, apres l'auoir longtemp travaillé, se creuva, et se vuida. |

Attention : parfois du texte peut sembler être ajouté alors que ce n'est qu'une « réclame »¹⁰. Nous donnons un exemple du projet Marillac (Marillac, chapitre 15, page 85) :



La dernière ligne « Ce qui a esté dict » est alors encodée ainsi :

¹⁰ Une réclame est placée au bas d'une page ; le ou les mots qui la composent correspondent au ou aux premiers mots de la page suivante.

```
<fw type="catch" place="margin-right"><del>Ce qui a esté dict</del></fw>
```

Précision : quand une longue partie de texte est-ajoutée – comme une page ou une partie de page, nous utilisons l'élément `<addSpan/>` pour éviter les problèmes de chevauchements. Il marque le début de la partie ajoutée, la fin de la partie étant marquée par l'élément cible `<anchor>`. Les attributs `@spanTo` et `@xml:id` permettent de marquer l'endroit du début et de la fin de l'ajout, `@spanTo` pointant vers l'identifiant de l'ancre.

Dans le chapitre 15 « Considerations » de l'exemplaire des Archives nationales, l'ajout de trois lignes page 2 verso peut être encodé ainsi :

```
<lb/><add>spirituelle, en sorte que j'ay veu telles fois</add>
<lb/><add>le Roy mesme <add>leur</add> demander moderation et
<expan>temperament</expan></add>
<lb/><add><del>de cette fervente affection</del> <add>en ceste maniere
d'affection</add> <del>et nonobstant ce</del></add>
```

Cet encodage n'est cependant pas satisfaisant car les lignes elles-mêmes sont ajoutées ce qui n'apparaît pas ici. Nous préconisons donc l'encodage suivant :

```
<addSpan/>
<lb/>spirituelle, en sorte que j'ay veu telles fois
<lb/>le Roy mesme <add>leur</add> demander moderation et
<expan>temperament</expan>
<lb/><del>de cette fervente affection</del> <add>en ceste maniere
d'affection</add> <del>et nonobstant ce</del>
<anchor xml:id="addEnd2v"/>
```

Ce procédé est également applicable aux ratures.

Enfin, il peut arriver que le texte de l'ajout soit sur une autre page. Nous donnons l'exemple des pages 50 verso et 51 du chapitre 13 du manuscrit Marillac : un ajout est signalé avec le caractère « + » dans un paragraphe de la page 50 verso et son texte ajouté à la page suivante.

```
[page 50 verso]Un capitaine des gardes+ Mesmes du temps qu'il estoit
Conseiller d'Estat...

[page 51]...m'a dict que luy portant parole de la part du Roy pour un
baron chargé à la verité de crimes peu remissibles ; il lui avoit
respondu en ces termes, je ne recois point commande de sa majesté que de
sa propre bouche.
```

qu'il faut lire ainsi :

```
Un capitaine des gardes m'a dict que luy portant parole de la part du Roy
pour un baron chargé à la verité de crimes peu remissibles ; il lui avoit
repondu en ces termes, je ne recois point commande de sa majesté que de
sa propre bouche. Mesmes du temps qu'il estoit Conseiller d'Estat...
```

Pour arriver à un tel résultat, nous utilisons l'élément `<ptr>` et l'attribut `@target` qui renvoie vers l'identifiant¹¹ :

```
Un capitaine des gardes+<add><ptr target="#add51r"/></add> Mesmes du
temps qu'il estoit Conseiller d'Estat<gap/><pb/><gap/>
<addSpan spanTo="#addEnd51" xml:id="add51r" rendition="+"/>m'a dict que
luy portant parole de la part du Roy pour un baron chargé à la verité de
crimes peu remissibles ; il lui avoit repondu en ces termes, je ne
recois point commande de sa majesté que de sa propre bouche
<anchor xml:id="addEnd51"/>
```

La feuille de style en XSLT permettra une présentation en HTML respectant la lecture intuitive du texte.

Ratures

L'élément `` est utilisé pour décrire les simples suppressions comme dans l'exemple tiré du chapitre des « mortifications » (Marillac, chapitre 11, page 314) :

```
Il me respondit, <del>Ils</del> Non, non, Il faict bon, s'accoustumer de
bonne heure de voir ces pourtraictz
```

En règle générale, nous marquons la suppression ligne par ligne, la ligne étant la structure minimale du texte comme dans l'exemple tiré du chapitre « Amys » (Marillac, chapitre 16, page 107) :

```
<lb/>s'estoient portez pour ses amys <del>dudict sieur de</del>
<lb/><del>Marillac</del> pendant qu'il estoit en charge n'ont
<lb/>point eu de stabilité pour luy en sa disgrace
```

Cette règle permet de contourner le problème des chevauchements en XML. Cependant, si aucun élément ne se chevauche et si la suppression concerne plusieurs lignes, l'élément `` peut englober l'ensemble des lignes comme dans l'exemple ci-dessous (Marillac, chapitre 11, page 319) :

```
<lb/><del>O bonne croix longtemps desirée, et preparée à celuy
<lb/>qui la recherche d'affection, Recevez en vous ce
<lb/>disciple de celuy qui a esté attaché.</del> Le merite,
```

¹¹ Pour cet exemple, nous utilisons l'élément `<gap/>` qui signifie la coupure d'un texte.

Quand une longue partie de texte est supprimée – comme une page ou une partie de page, nous utilisons l'élément `<delSpan>` pour éviter les chevauchements. `<delSpan>` marque le début de la partie supprimée, la fin de la partie étant marquée par l'élément cible `<anchor>`. Les attributs `@spanTo` et `@xml:id` permettent de marquer l'endroit du début et de la fin de la suppression, `@spanTo` pointant vers l'identifiant de l'ancre.

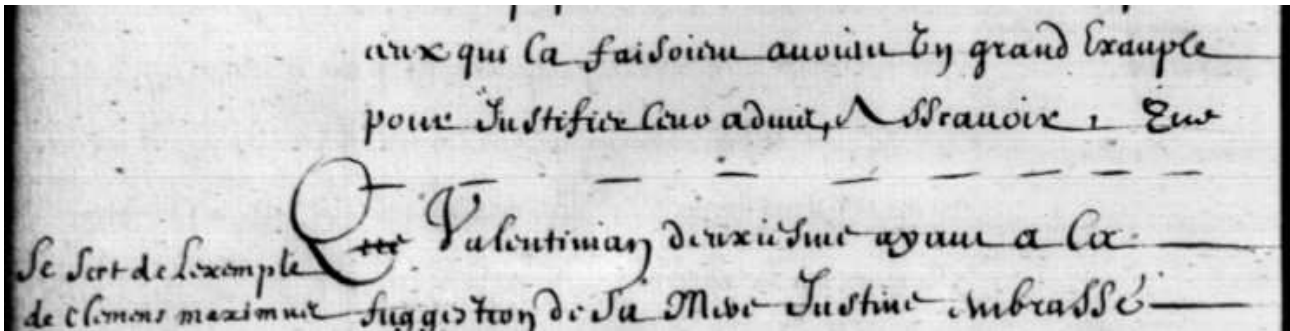
Reprenons notre dernier exemple ; nous proposons un nouvel encodage :

```
<lb/><delSpan spanTo="#delEnd319r"/>O bonne croix longtemps desirée, et
preparée à celuy
<lb/>qui la recherche d'affection, Recevez en vous ce
<lb/>disciple de celuy qui a esté attaché,<anchor xml:id="delEnd319r"/>
le merite,
```

Pour l'attribut `@xml:id`, nous proposons les valeurs suivantes :

- `delEnd319r` la rature se termine à la fin de la page 319 recto
- `delP319r` la rature se termine à la fin du paragraphe de la page 319 recto
- `addTop6r` l'ajout se termine au début de la page 6 recto

Parfois, une structure comme celle du paragraphe est supprimée. Nous utilisons alors également l'élément `<delSpan>` comme dans l'exemple ci-dessous (Marillac, chapitre 3, page 97 verso) :



Ici, l'auteur relie finalement deux paragraphes par le biais d'un trait pointillé, passage que nous encodons ainsi :

```
<gap/>
<lb/>pour justifier leur avis, asscavoir. <add>que</add>
<delSpan spanTo="#delEnd97v" rend="dash"/>
</p>
<p>
<anchor xml:id="delEnd97v"/><del>Que</del>
<lb/>Valentinian deuxiesme ayant à la
<gap/>
```

Le trait discontinu est encodé par la valeur "dash" de l'attribut `@rend`.

Le rendu

Pour l'élément ``, par défaut, la feuille de style prévoit un simple trait. Cependant certaines ratures sont plus complexes : nous utilisons alors l'attribut `@rend` pour donner le rendu de la suppression sur la source originale avec les valeurs suivantes :

| | |
|--|---|
| <code>"strikethrough"</code> | simple trait |
| <code>"strikeCircle"</code> | pour les ronds circulaires |
| <code>"strikeDiagonalLR"¹²</code> | le trait diagonal commence au-dessus de la ligne à gauche et se poursuit vers la droite |
| <code>"strikeDiagonalRL"</code> | le trait diagonal commence au-dessus de la ligne à droite et se poursuit vers la gauche |
| <code>"strikeCross"</code> | les passages sont rayés par une croix |
| <code>"strikeVertical"</code> | trait vertical |
| <code>"strikeVerticaldouble"</code> | double traits verticaux en parallèle |
| <code>"strikeDiagonalLRdouble"</code> | double traits en diagonal de haut en bas de gauche à droite en parallèle |
| <code>"strikeDiagonalRLdouble"</code> | double traits en diagonal de haut en bas de droite à gauche en parallèle |

Le problème reste l'affichage de ces traits verticaux ou obliques qui sera résolu avec les feuilles de style CSS et XSLT.

Remarque : parfois le texte raturé est rendu illisible. Ici nous ne suivons pas les Guidelines qui recommandent l'élément `<gap>` ; dans ce texte, cette illisibilité est voulue par les scribes eux-mêmes : nous utilisons la valeur « `strikeCircle` » pour montrer leur volonté de rendre des parties illisibles.

Suppressions et ajouts concomitants

Fréquemment les ajouts et les suppressions sont concomitants dans le manuscrit Marillac (Marillac, chapitre 16, page 99 verso) :

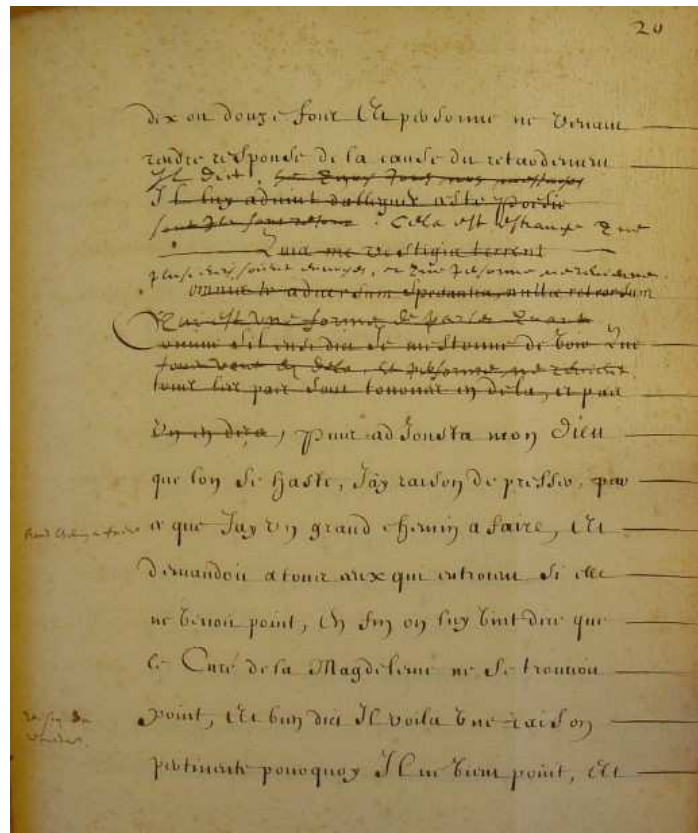
```
<p>
<lb/>Il me disoit souvent ce proverbe <del>Pacifici tibi sunt</del>
<add>multi pacifici sint tibi</add>
<lb/><del>multi sed con consiliarius unus de mille,</del> <add>et
consiliarius sit tibi unus de mille.</add>
</p>
```

Remarque : pour privilégier la fluidité de la lecture et assurer une meilleure lisibilité en html, nous avons décidé de mettre un espace entre les `` et les `<add>`.

Parfois les ajouts sont eux-mêmes supprimés. Il convient de bien respecter la structure (la ligne) et le processus de l'écriture au détriment de la mise en forme réelle dans le manuscrit. Prenons

¹² Les lettres « L » et « R » signifient « left » et « right ».

l'exemple du chapitre Décès des Archives nationales ; voici le passage concerné tel qu'il apparaît dans le manuscrit (Marillac, chapitre 21, page 20) :



et tel que nous l'encodons :

```
<lb/><del>Il luy advint d'alleguer ceste poesie</del> <add>Il dict
<del>et que pour nous messieurs</del></add>
<lb/><del>Quia me vestigia terrent</del> <add><del> sont plus sans
response</del> cela est estrange que</add>
<lb/><del>Omnia te adversum spectantia, nulla retror sum</del>
<add>plusieurs soient envoyés, et que personne ne revienne.</add>
<lb/><del>Comme s'il eust dict je m'estonne de voir que</del>
<add><del>Qui est une forme de parler quand</del></add>
<lb/><del>tous les pas sont tousjours en delà et pas</del> <add><del>tous
vont en dela, et personne ne revient,</del></add>
<lb/><del>un en deça</del>.
```

Voici maintenant le même passage tel que nous le restituons dans sa version html :

[20 r]

dix ou douze fois et personne ne venant

rendre response de la cause du retardement

~~Il luy advint d'alleguer ceste poesie Il dict et que pour nous messieurs~~~~Quia me vestigia terrent sont plus sans response cela est estrange que~~~~Omnia te adversum spectantia, nulla retror sum~~ plusieurs soient envoyés, et que personne ne revienne.~~Comme s'il eust dict je m'estonne de voir que Qui est une forme de parler quand tous les pas sont toujours en deçà et pas tous vont en deçà, et personne ne revient,~~~~un en deça.~~ Puis adjousta : mon Dieu

que l'on se haste. J'ay raison de presser, par-

ce que j'ay un grand chemin à faire. Et

grand chemin à faire

demandoit à tous ceux qui entroient si elle

ne venoit point. Enfin on luy vint dire que

le curé de la Magdeleine ne se trouvoit

point. Et bien dict-il voila une raison

raison du retardement

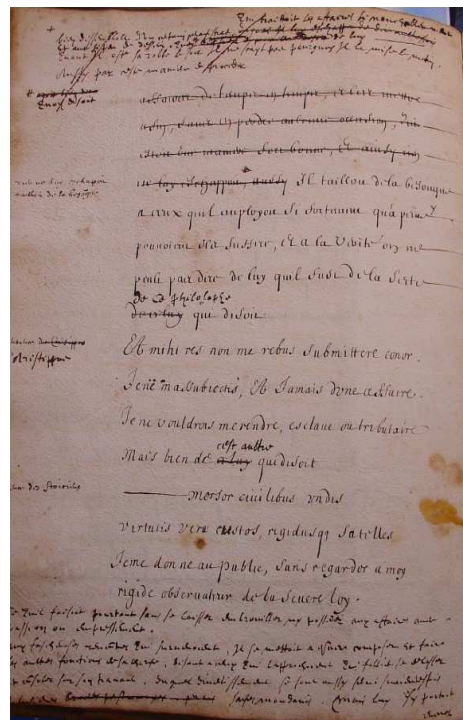
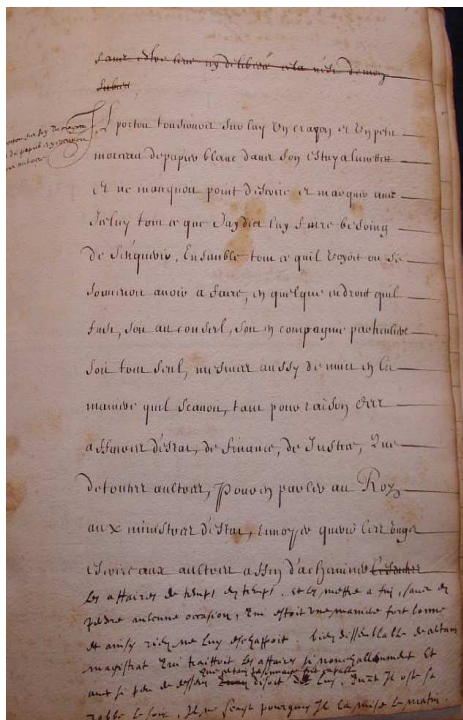
pertinente pourquoy Il ne vient point, et

Constatons que les ajouts sont sur la ligne à la suite des passages raturés, ce qui rend plus lisible le texte mais au détriment de l'aspect diplomatique de la source.

Avec <delSpan> et <addSpan>

L'encodage se complique parfois quand le processus d'écriture est difficilement compréhensible en raison du nombre de ratures et d'ajouts.

Prenons l'exemple du chapitre « Capacités » des Archives nationales ; un paragraphe a été ajouté en fin de page 5 recto (Marillac, chapitre 12, page 5). Cependant, à la lecture de la page suivante, la page 5 verso, le processus d'écriture apparaît : le paragraphe a été préalablement ajouté sur cette page avant d'être supprimé puis ajouté sur la page précédente.



Il est possible de montrer le séquençage des ratures et des ajouts. L'attribut @seq est alors requis.

Nous proposons l'encodage suivant des deux ajouts :

```

<lb/>escrire aux aultres affin d'acheminer <del>lesdictes</del>
<addSpan spanTo="#addEnd5r" seq="2"/>
<lb/>Les affaires de temps en temps et les mettre à fin sans en
<lb/>perdre aulcune occasion, qui estoit une maniere fort bonne
<lb/>et ainsy rien ne luy eschappoit, bien dissemblable d'un certain
<lb/>magistrat qui traittoit les affaires si nonchallement et
<lb/>avec si peu de dessein que certain personnage fort capable ? disoit
de luy quand il oste sa
<lb/>robbe le soir il ne scayt pourquoy il l'a mise le matin.
<anchor xml:id="addEnd5r"/>
<pb/>
<fw type="pageNum">[5 v]</fw>
<addSpan spanTo="#addTop5v" seq="1"/>
<delSpan spanTo="#delTop5v" rend="strikeDiagonal-rightleft"/>
<lb/>+ bien dissemblable d'un certain magistrat <del>duquel il luy
eschappa de dire aultresfois</del> <add>qui traittoit les affaires si
nonchallement</add>
<lb/><add>et avec si peu de dessein que ++ <add place="margin-left">
<del>esprit bien <unclear/></del> <add>quon disoit</add> </add>certain
personnage fort capable disoit de luy</add> Quant il oste sa robbe le
soir il ne scayt pas pourquoy il l'a mise le matin
<lb/>Aussi par ceste manière de proceder
<anchor xml:id="delTop5v"/>
<anchor xml:id="addTop5v"/>

```

Dans les choix opérés pour cet encodage du manuscrit Marillac, il manque l'élément `<subst>` pour rendre compte de la succession des natures différentes des ratures et ajouts qui pourrait être :

```

<subst>
<del seq="1" rend="strikethrough">Monsieur</del> <add
seq="1"><del seq="2" rend="strikeCircle">M.</del></add>
</subst>

```

Cependant, nous avons décidé de ne pas rendre la finesse du processus d'écriture dans la première phase du projet. De plus, les éléments `` et `<add>` encodent formellement le texte alors que `<subst>` est un élément sémantique : c'est déjà une première analyse du texte. Nous ne mélangeons pas les deux types d'encodage ici¹³. En revanche, nous pouvons utiliser exceptionnellement l'attribut `@seq`.

¹³ Le TEI concil se pose d'ailleurs la question de retirer l'élément `<subst>`.

11 Numérotation multiple des pages et des feuillets

Quand les numéros de pages et de feuillets sont multiples, nous ne les considérons pas comme des ajouts. Nous utilisons l'élément `<fw>` pour les numérotations de pages et quand nous avons deux numéros différents, l'un à gauche et l'autre à droite, nous doublons l'élément `<fw>`, ainsi :

```
<fw type="pageNum" place="top-left">7</fw>
<fw type="pageNum" place="top-right">5</fw>
```

Parfois, l'un des deux peut être rayé et nous ajoutons alors l'élément ``:

```
<fw type="pageNum" place="top-right"><del>20</del> <add>319</add></fw>
```

Le balisage sémantique

A la différence de l'analyse lexicale qui s'occupe d'un mot dans une phrase, l'analyse sémantique s'intéresse à la signification de ce mot.

Le balisage sémantique autorise un encodage d'un premier niveau d'analyse comme les langues, les références bibliographiques, les paroles et les écrits rapportés, et les entités nommées.

Dans le projet Marillac, il a été décidé d'encoder sémantiquement uniquement l'exemplaire de la Bibliothèque Sainte Geneviève, notre exemplaire témoin.

1 Les langues

La langue principale des textes encodés dans les deux projets est le français. Des phrases et citations sont en latin et en grec ; quelques phrases en italien. Par ailleurs, nous distinguons le grec ancien écrit en alphabet latin du grec ancien écrit en alphabet grec.

Exceptionnellement, nous ne suivons pas complètement les recommandations du consortium TEI énoncées dans les Guidelines. En effet, ces dernières préconisent de se conformer au standard BCP 47 – code de langue IETF – ; or comme les deux projets ne présentent pas de problème particulier lié à l'utilisation des langues, nous choisissons la simplicité, c'est-à-dire utiliser la norme ISO 639, celle-ci étant, au demeurant, une composante du standard BCP¹⁴.

Dans l'en-tête TEI, nous identifions les langues selon la norme ISO 639-3 qui définit les langues avec un niveau de détail linguistique :

```
<langUsage>
  <language ident="fre">Français</language>
  <language ident="lat">Latin</language>
  <language ident="grc">Grec ancien avant 1453</language>
  <language ident="gss">Grec ancien en alphabet grec</language>
  <language ident="ita">Italien</language>
</langUsage>
```

¹⁴ La norme BCP comprend des sous-étiquettes : l'une d'entre elle est issue de la norme ISO 639.

Dans l'élément `<text>`, on utilise l'attribut `@xml:lang` avec la valeur `"fre"` pour indiquer la langue principale, ici la langue française :

```
<text xml:lang="fre">
```

Enfin, dans le corps du texte, il y a deux possibilités de marquer les langues secondaires avec l'attribut `@xml:lang` :

1. Le texte n'est pas balisé : on utilise l'élément `<foreign>` dans lequel on met l'attribut `@xml:lang` avec sa valeur :

```
<foreign xml:lang="gss">ΚΑΤέΠιgέvέσίv</foreign>
```

2. Le texte est balisé : on met l'attribut `@xml:lang` avec sa valeur dans l'élément :

```
<quote xml:lang="lat">dimitte me paululum ut glutiam salivam  
meam</quote>
```

Dans la pratique, nous privilégions l'utilisation de l'élément `<foreign>` même si un autre élément comme `<quote>` ou `<p>` nous permettrait d'en faire l'économie. En effet, d'une part l'attribut `@xml:lang` exige la déclaration dans la feuille de style de tous les éléments concernés et, d'autre part, l'élément `<foreign>` permet de contourner tous les problèmes liés aux chevauchements. Enfin, il n'est pas recommandé de mélanger les balisages formel et sémantique.

2 Les références bibliographiques

Les références bibliographiques comprennent au moins un titre d'ouvrage, de chapitre, ou tout autre partie d'un texte (préface, éloge, épître, etc.). Peuvent s'y ajouter l'auteur, la date, le numéro de page, l'éditeur et l'imprimeur des textes publiés.

Le projet Patin, en particulier, comporte beaucoup de références bibliographiques qui nécessitent un encodage précis. Ceci contribue d'ailleurs à la singularité de ce projet. : l'encodage permettra à terme de constituer un catalogue bibliographique de la bibliothèque utilisée par l'auteur.

Pour référencer bibliographiquement un texte, nous utilisons les éléments suivants : `<bibl>`, `<title>` et `<author>` ; nous avons choisi de ne pas structurer les autres références : éditeur, imprimeur, lieu, etc. En effet, les références y sont incomplètes, par exemple (Patin, page 1) :

```
<bibl>  
<title>la Republique</title> de <author>Bodin</author>  
</bibl>
```

Parfois, l'élément `<title>` est utilisé sans son parent `<bibl>` pour un titre de discours par

exemple ou tout autre intitulé de texte qui indique son contenu sans pour autant être un titre de livre ou d'article :

```
un discours intitulé, <title type="speech">Suasion de l'arrest au
parlement pour la manutention de la loy salique</title> en l'argument
duquel il a mis
```

C'est aussi le cas du « Traité de la vie éternelle », manuscrit écrit par Michel de Marillac et jamais imprimé.

Enfin, nous typons les différentes références bibliographiques comme les libelles (Patin, page 77) :

```
Aussi est-il auteur du libelle intitulé <bibl
type="libelle"><title>Admonitio ad Regem Christianissimum</title></bibl>
```

Se pose la question de régulariser les titres qui sont parfois donnés dans une forme raccourcie ou de façon farfelue. Dans le cadre du projet Patin, cette régularisation est envisagée dans une seconde phase.

3 Discours et citations¹⁵

Nous tenons à remercier Dora de Lima qui a accepté de travailler de façon collaborative sur cet aspect précis de l'encodage : elle nous a beaucoup apporté.

Dans le texte écrit, il est nécessaire de distinguer visuellement les différentes structures de la parole rapportée, de la pensée et des écrits. Dans les règles typographiques actuelles, le discours oral se repère par des guillemets, les titres de livres par le style italique, etc. Les données structurées reprennent ces règles.

Pour encoder les discours et les citations, nous distinguons deux types :

- 1) les paroles rapportées ou « speech » en anglais, c'est-à-dire les discours oraux, directs ou indirects ;
- 2) les écrits rapportés mot pour mot soit sans références précisées soit avec références bibliographiques : nous nommons ces écrits « textes ».

Quand ces textes sont transcrits exhaustivement, ils deviennent des documents (lettres, missives, arrêts, etc.) et sont encodés différemment.

Les discours oraux

Pour les discours oraux, directs et indirects, nous utilisons l'élément `<q>`¹⁶. L'attribut `@type` avec la valeur « direct » précise le discours direct parlé et l'attribut `@type` avec la valeur « indirect » le discours indirect. Nous donnons deux exemples, un pour chaque type.

1. Discours direct

¹⁵ Pour Dora De Lima, les éléments `<bibl>` et `<titl>` correspondent, dans la typologie d'Antoine Compagnon, au « fait de donner le titre d'une œuvre et/ou le nom d'un auteur auxquels on renvoie, qui accompagnent, ou non, une citation ». (Antoine COMPAGNON, *La seconde main ou le travail de la citation*, Paris, Seuil, 1979, p. 59.).

¹⁶ Lou Burnard nous a conseillées de créer un élément spécifique à notre projet que nous pourrions appeler : `<speech>`. Nous le préconisons dans la poursuite et seconde phase du projet.

il luy respondit je me suis trouvé mal mais cela n'est rien. (Marillac, chapitre Décès, page 224 v)

que nous encodons ainsi :

```
il luy respondit <q type="direct">je me suis trouvé mal mais cela n'est rien.</q>
```

Quand le locuteur est explicitement indiqué par l'expression « dict-il » ou « respondit-il » par exemple, nous ne l'encodons pas expressément (Marillac, chap. 21) :

```
<q type="direct">Dieu soit loué, respondit il, j'en suis bien ayse, j'espere que Dieu aura pitié de son église et qu'il détruira ceux qui la persecutent.</q>
```

Ici le locuteur est Michel de Marillac.

2. Discours indirect

Et ceste austerité et mortification alla si avant qu'il tomba en une grande maladie pour laquelle une certaine personne qui luy estoit fort confidente obtint de luy qu'il se desisteroit de ceste austerité, ainsy que depuis elle m'a dict. (Marillac, chapitre Mortifications, page 300 recto)

que nous encodons ainsi :

```
Et ceste austerité et mortification alla si avant qu'il tomba en une grande maladie pour laquelle une certaine personne qui luy estoit fort confidente <q type="indirect"> obtint de luy qu'il se desisteroit de ceste austerité</q> ainsy que depuis elle m'a dict.
```

Les exemples ne sont pas toujours aussi simples. Dans les textes, le discours oral peut provenir d'une citation extérieure au texte, empruntée – qu'elle soit explicite ou implicite – comme un proverbe, un passage d'un livre, des vers, etc.. Nous pouvons l'indiquer par l'attribut @type avec la valeur « written »¹⁷ :

```
<q type="indirect written">blablabla</q>
```

Pour spécifier le locuteur, nous ajoutons l'attribut @who ; il indique le personnage ou le groupe de personnes à qui l'on attribue le discours grâce à une valeur qui pointe vers un identifiant.

Reprenons l'exemple du discours indirect que nous complétons :

¹⁷ A ce stade du projet, cette valeur n'est pas encore utilisée.

Et ceste austerité et mortification alla si avant qu'il tomba en une grande maladie pour laquelle une certaine personne qui luy estoit fort confidente <q type="indirect" who="#MDBFM"> obtint de luy qu'il se desisteroit de ceste austerité</q> ainsy que depuis elle m'a dict.

La valeur "#MDBFM" correspond à Madeleine Du Bois de Fontaine-Marans dite Madeleine de Saint-Joseph nommée souvent dans le texte « personne fort confidente ».

Dans cette première version, seul le locuteur est spécifié ; dans une seconde phase, il est d'ores et déjà prévu de préciser à quel personnage s'adresse le discours.

Dans un fichier externalisé, nous dressons la liste des intervenants grâce à l'attribut @xml:id et aux éléments suivants : <listPerson>, <personGrp>, <person>, <persName>, <name>, <state>, <label>, <desc>, <name> et <p>¹⁸.

Reprenons notre exemple précédent. L'attribut @who pointe vers l'identifiant "MDBFM" ; celui-ci, créé par nos soins, correspond aux initiales du nom de la personne. Grâce à l'élément <name>, nous ajoutons « personne fort confidente » qui identifie ce personnage :

```
<listPerson>
  <person xml:id="MDBFM">
    <persName>Madeleine Du Bois de Fontaine-Marans</persName>
    <state>
      <label>Carmélite</label>
      <desc>
        <persName>
          Madeleine de Saint-Joseph
          <name>personne fort confidente</name>
        </persName>
      </desc>
    </state>
  </person>
  <person>...</person>
</listPerson>
```

Reste à déterminer l'utilisation de toutes ces appellations et comment elles apparaîtront sur le site dans des index différents à préciser.

Les citations écrites

Par citation, nous entendons un « Passage emprunté à un auteur qui peut faire autorité » (Littré) qu'il soit explicite ou implicite.

Les citations extérieures au texte sont de deux types : les premières sans références précisées, et les secondes avec références bibliographiques¹⁹.

18 Tous ces éléments peuvent être déclarés dans l'en-tête grâce à l'élément <particDesc> dans <teiHeader> sans pour autant faire l'objet d'un fichier externalisé. Ici, le nombre important de personnages et de chapitres nous a convaincu d'utiliser cette possibilité.

19 Dora De Lima nous a rapporté que pour Antoine Compagnon, la citation est considérée comme la procédure minimale d'insertion d'un texte dans un autre, c'est-à-dire comme une « relation interdiscursive primitive ». Les deux éléments <quote> et <cit> correspondent aux formes de l' « hétérogénéité montrée » du récit. Voir. Jacqueline Authier-Revuz, « Hétérogénéité(s) énonciative(s) », *Langages*, n° 73, mars 1984, p. 84-111.

- Pour les premières, nous utilisons l'élément `<quote>` seul.

Il y a aussy quelques aultres natures d'afflictions qu'un autheur de nostre temps appelle essay à fer emoulu...
(Marillac, chapitre 11, page 318)

que nous encodons ainsi :

Il y a aussy quelques aultres natures d'afflictions qu'un autheur de
nostre temps appelle `<quote>`essay à fer emoulu`</quote>`,

Quand le texte est en vers, l'élément `<quote>` est associé aux deux éléments `<lg>` et `<l>`, ainsi (Marillac, chapitre 1, page 22 [44 verso]) :

```
<quote type="epitaphe">
  <lg>
    <l>Cy gist un grand prelat, sans reproche,et sans blasme,</l>
    <l>de l'eglise en son temps, l'ornement, et maintien,</l>
    <l>qui par son testament, donna a Dieu son ame,</l>
    <l>a la terre son corps, et aux pauvres son bien.</l>
  </lg>
</quote>
```

Parfois, ces citations sont tirés de sources comme des lettres, des missives ou des libelles : ce sont de simples extraits qui ne correspondent pas à l'intégralité du document²⁰. La citation est alors typée comme dans l'exemple suivant (Marillac, chapitre 11, page 318 verso) :

```
<lb/><quote type="letter">Il a pleu à Dieu que ma vie depuis le
commencement
<lb/>jusques à ceste heure, n'a esté que parmy les croix
<lb/>en telle sorte, que les grandeurs, et prosperitez mesmes
<lb/>ont esté plus ameres, que les mortz de tous mes
<lb/>proches, Je finiray volontiers avec mon maistre,
<lb/>et expireray sur la croix, qui me sera grande
<lb/>faveur, et consolation. et en quelque sorte qu'il
<lb/>arrive, j'attends bien ceste suite de la providence
<lb/>divine sur moy, qu'il ne permettra point que je
<lb/>soit separé de la croix</quote>
```

Puis nous encodons l'auteur avec l'élément `<ptr>` et l'attribut `@target` qui pointe sur l'identifiant de la personne²¹ indiqué dans le fichier externalisé de la liste des personnages.

20 Si la citation reprend l'intégralité du document, l'élément `<floatingText>` est alors requis. Pour Dora De Lima, cet élément permet de mettre en évidence le procédé d'enchâssement. « La différence avec d'autres pratiques intertextuelles comme la citation ou le pastiche par exemple, est qu'ici les deux textes sont dans un rapport d'enchâssement et de réflexion tout à la fois : le récit enchâssé se doit de refléter le contenu fictionnel, métatextuel ou transcendantal du récit enchâssant. » Lucien Dällenbach, *Le récit spéculaire. Essai sur la mise en abyme*, Paris, Seuil, 1977, p. 76.

21 Pour les références croisées, voir la partie ci-après.

Enfin, dans le pointeur, l'attribut @type avec la valeur « author » est ajouté²² :

```
<lb/><quote type="letter">Il a pleu à Dieu que ma vie depuis le
commencement
<lb/>jusques à ceste heure, n'a esté que parmy les croix
<lb/>en telle sorte, que les grandeurs, et prosperitez mesmes
<lb/>ont esté plus ameres, que les mortz de tous mes
<lb/>proches, Je finiray volontiers avec mon maistre,
<lb/>et expireray sur la croix, qui me sera grande
<lb/>faveur, et consolation. et en quelque sorte qu'il
<lb/>arrive, j'attends bien ceste suite de la providence
<lb/>divine sur moy, qu'il ne permettra point que je
<lb/>soit separé de la croix<ptr target="#MDMA" type="author"/></quote>
```

- Pour le seconde type de citations avec références bibliographiques, l'élément <cit> est recommandé. Même partielles, les références sont alors encodées avec les éléments de structuration bibliographique minimale : <bibl> et ses enfants <title> et <author>.

```
comme dit Erasme in Ciceroniano rarum fuit nature miraculum (Patin, page
16)
```

que nous encodons ainsi :

```
comme dit
<cit>
<bibl><author>Erasme</author> <foreign xml:lang="lat">in</foreign> <title
xml:lang="lat">Ciceroniano</title></bibl>
<quote xml:lang="lat">rarum fuit nature miraculum</quote>
</cit>
```

Enfin, le nom de l'auteur peut être donné dans la citation même, explicitement ou implicitement. Nous encodons alors la citation avec l'élément <ref> pour marquer l'auteur du texte (Marillac, Chapitre 11, page 321) :

```
<cit>
<quote>la pourpre, l'applaudissement, et la puissance, qui sont <ref
target="marillac_listbiblio.xml#seneque" type="author"/>dict
Seneque</ref> les trois pointcs où l'ambition tend ;</quote>
</cit>
```

ou encore (Marillac, Chapitre 6, page 171) :

²² Cette possibilité nous a été suggérée par Lou Burnard ; qu'il en soit ici remercié.

```
Cela est conforme à ce qui est escript dans <cit><ref
target="marillac_listbiblio.xml#SIRACIDE"><title>l'Ecclesiastique</title><
/ref> <quote><foreign xml:lang="lat">Dives locutus est et omnes tacuerunt
et verbum illius usque ad nube perducent, pauper locutus est sensate et
non est datus et locus at dicunt quis es hic.</foreign></quote></cit>
```

D'autres cas sont plus compliqués à encoder. Si la citation avec la référence bibliographique n'est pas structurée, les éléments `<cit>` et `<bibl>` ne sont plus utilisés au profit uniquement des éléments `<title>`, `<quote>` et `<ref>`, ce dernier élément renvoyant la référence dans une liste bibliographique externalisée :

```
Aussy Michel de Montaigne en son Essay disoit toutes aultres choses tombent en commerce
nous prestons nos biens et nos vies au besoing de nos amys mais de communiquer nostre
honneur et d'estonner aultruy de nostre gloire il ne se void guere.
```

extrait encodé ainsi :

```
Aussy <persName>Michel de Montaigne</persName> en son <ref
target="#MONTAIGNE"><title>essay</title></ref> disoit
<quote>toutes aultres choses tombent en commerce nous
prestons nos biens et nos vies au besoing de nos amys mais de
communiquer nostre honneur et d'estonner aultruy de nostre
gloire Il ne se void guere<ptr target="#MONTAIGNE"/></quote>.
```

Nous constatons une répétition de la cible « MONTAIGNE » en tant qu'auteur de l'ouvrage mais aussi de la citation :

```
<ref target="#MONTAIGNE"><title>essay</title></ref>
```

et

```
<ptr target="#MONTAIGNE"/>
```

les éléments `<ref>` et `<ptr>` renvoyant à la référence bibliographique, le premier pour le titre, le second pour l'auteur de la citation. Cette répétition permet la construction de deux index différents : une bibliographie d'une part et une liste d'auteurs et de leurs écrits d'autre part.

Ici, utiliser l'élément `<cit>` n'est pas valide en raison du texte libre comme « en son ». Quand la référence bibliographique est noyée dans la narration, il n'est donc pas possible de structurer efficacement les données. Il est cependant possible de ne pas respecter strictement les recommandations de la TEI et d'utiliser l'élément `<bibl>` dans `<cit>` comme nous le faisons régulièrement dans le projet Patin.

4 Les entités nommés

Définition : les entités nommés sont « Tous les éléments du langage définis par référence : les noms propres au sens classique, les noms propres dans un sens élargi mais aussi les expressions de temps et de quantité » (Friburger, 2006)²³.

La temporalité

Dans le manuscrit, le temps est rythmé par deux éléments distincts :

- `<date>` pour une date du calendrier
- `<time>` pour l'heure

Le premier a pour attribut `@when` quand la date est précise, par exemple :

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

la valeur de l'attribut `@when` se déclinant ainsi (à l'anglo-saxonne) : `"aaaa-mm-jj"`.

Le second est déjà plus difficile à appréhender. Nous avons choisi d'utiliser l'élément `<time>` avec les attributs `@when`, `@from`, `@to`, `@notBefore` et `@notAfter` :

- `@when` pour une heure précise :

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

- `@from` et `@to` pour un intervalle de temps certain :

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

- `@notBefore` et `@notAfter` pour un intervalle incertain :

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

Timeline²⁴

Nous construisons une frise chronologique ou ligne de temps selon les modèles du site « simile-widgets » (voir le site www.simile-widgets.org avec les exemples).²⁵

²³Friburger (2006) « Linguistique et reconnaissance automatique des noms propres », revue Méta, *la traduction des noms propres et Langue, traduction et mondialisation : interactions d'hier, interactions d'aujourd'hui*, vol.51, n°4, p. 637-650, décembre 2006, cité dans « Extraction des chemins entre deux entites nommees en vue de l'acquisition des patrons 'de relations » Yayoi Nakamura-Delloye ALPAGE, INRIA, disponible en ligne http://hal.archives-ouvertes.fr/docs/00/51/14/81/PDF/P120_Poster62.pdf

²⁴ L'élément de la TEI `<timeline>` n'est pas ici utilisé car il s'applique à des textes oraux.

²⁵ L'idée de la Timeline est venue à Valérie Gratsac-Legendre au fur et à mesure de l'encodage, idée qu'à mis en œuvre

Nous nous appuyons d'une part sur la temporalité marquée par les éléments <date> et <time>, et d'autre part sur les notes marginales. Parfois au détour d'une phrase, nous découvrons un événement important, par exemple : « ...il estoit arrivé à Paris avant le Roy au retour de la prise de la Rochelle... ». Ce contexte historique est alors marqué dans la Timeline bien que la date exacte ne figure pas explicitement dans le texte : elle est identifiable en tant que telle.

Les éléments requis sont au nombre de cinq avec sept attributs :

1. <listEvent> est l'élément parent
2. <event> a pour enfant les trois éléments suivants :
 1. <head>
 2. <label> a pour correspondance une icône
 3. <desc>

et les attributs suivants :

1. @xml:id,
2. @when,
3. @notBefore,
4. @notAfter,
5. @from,
6. @to,
7. @where.

Voici quelques exemples tirés du chapitre « Décès » du manuscrit Marillac (Marillac, chapitre 22) :

```
<listEvent>
  <event xml:id="deces-01" from="1632-08-01" to="1632-08-"
  where="chateaudun">
    <head>Décès de Marillac</head>
    <label>Chapitre</label>
  </event>
</listEvent>
```

Puis :

```
<listEvent>
  <event xml:id="deces-02" when="1632-08-01"
  notBefore="23:00:00" notAfter="24:00:00">
    <head>Début de la maladie</head>
    <label>Médecine</label>
```

```

    </event>
</listEvent>

```

Ou encore :

```

<listEvent>
  <event xml:id="deces-06" when="1628-10-28">
    <head>Prise de La Rochelle</head>
    <label>Contexte</label>
    <desc>La prise de La Rochelle, sous Louis XIII, a lieu le 28
octobre 1628.</desc>
  </event>
</listEvent>

```

- L'élément `<listEvent>` est situé au début de paragraphe et peut comprendre plusieurs `<event>`. Ici, la puce sera positionnée sur la date du « 28-10-1628 », en cliquant dessus, l'internaute naviguera sur le paragraphe.
- L'élément `<event>` comprend obligatoirement l'attribut `@xml:id`. Celui-ci reprend le nom du chapitre et un ordre numérique. S'y ajoute l'attribut `@when` où est saisie la date régularisée. Si l'heure est indiquée dans le texte, les attributs `@notBefore` et `@notAfter` sont requis comme dans notre exemple. Pour signaler la durée de l'événement, on utilise les attributs `@from` et `@to` qui donnent l'intervalle.
- L'élément `<head>` balise le titre de l'événement qui apparaîtra sur la ligne de temps sur lequel l'internaute pourra cliquer.
- Enfin, l'élément `<label>` visualisera par la couleur le type de l'événement par exemple le chapitre en rouge, le service public en bleu.

Pour le projet Marillac, ils sont au nombre de cinq, sont suggestifs et peuvent être modifiés au fur et à mesure du projet :

1. Chapitre
2. Médecine
3. Religion
4. Service public
5. Contexte

Pour le projet Patin, nous retenons trois autres types :

1. Section
2. Contexte
3. Publication

En effet, le texte comporte une liste de dates de publication de livres impressionnante, dates qu'il nous semble intéressantes de marquer temporellement.

Cette ligne de temps est affichée chapitre par chapitre. Cependant, nous préconisons qu'elle le soit sur l'ensemble des chapitres afin de permettre une navigation et une exploration différente du corpus.

Les noms de personnes et de lieux

Nous établissons différents index de noms de personnes et de lieux. Pour les dresser, plusieurs éléments sont utilisés selon des marquages précis.

Les personnages

Selon les Guidelines, les individus sont classés soit par leur identité soit sous forme d'un syntagme nominal, par exemple « conseiller à la cour » qui est une fonction ou « surintendant des finances » qui est une charge.

Historiquement, la TEI proposait l'élément `<rs>` pour *referencing string* afin de baliser les chaînes de caractères comprenant entre autres les entités nommés, ce qui donnait : `<rs type="name">`. Puis, cet encodage a été remplacé par l'élément `<name>` et enfin depuis la version P4 par `<persName>`.

- `<persName>` : « contient un nom propre ou une expression nominale se référant à une personne, pouvant inclure tout ou partie de ses prénoms, noms de famille, titres honorifiques, noms ajoutés, etc. ».

Nous donnons plusieurs exemples :

```
<persName>M. de Marillac</persName>
```

```
<persName>M. le Garde des sceaux du Vair</persName>
```

```
<persName>Roy de Navarre</persName>
```

```
<persName>archevêque de Lyon</persName>
```

- `<roleName>` : « contient un composant du nom d'une personne, indiquant que celle-ci a un rôle ou une position particulière dans la société, comme un titre ou un rang officiel. »
 1. `<roleName>` est imbriqué dans `<persName>` quand le personnage est identifié par son titre, par exemple :

```
<persName>Henri III <roleName>roi de France</roleName></persName>
```

Dans ces cas là, nous n'utilisons pas `<roleName>` et ne gardons que `<persName>`. Le résultat de l'encodage est donc le suivant :

```
<persName>Henri III roi de France</persName>
```

2. `<roleName>` est utilisé seul sans `<persName>` quand le titre ou une fonction existe indépendamment de la personne qui le porte par exemple :

```
Les <roleName>conseillers en la Cour de parlement</roleName>
```

Nous choisissons ici de ne pas dresser une typologie qui permettrait pourtant de distinguer les titres des fonctions. Dans la première version du projet, cet encodage est en suspens.

- <orgName> : « contient le nom d'une organisation. » par exemple « hostel de ville de Paris ».

Nous n'encoderions pas de la même manière « hostel de ville de Paris » et « hostel de ville à Paris », à savoir :

```
<orgName>hostel de ville de Paris </orgName>
```

et

```
<orgName>hostel de ville</orgName> à <placeName>Paris</placeName>
```

Dans la première version du projet, cet encodage est également en suspens.

Parfois, l'écriture est plus complexe et l'encodage semble bien difficile à réaliser. Par exemple, le nom peut être détaché de sa fonction et de l'institution pour mettre en valeur ses dernières. Nous privilégions alors les éléments <orgName> et <roleName>. Voici un exemple avec une solution que nous proposons :

```
Le <date when="1593-06-23">vingt troisieme jour de juin 1593</date>,
<del>Monsieur</del> <persName><add>M.</add> de Marillac</persName> lors
<roleName>conseiller en la <orgName>cour de parlement</orgName> en la
<orgName>seconde Chambre des enquestes</orgName></roleName>, estant au
<orgName>bureau</orgName> de ladicte chambre, representa à ladicte
Compagnie, que l'on traittoit en l'assemblée des <orgName>estatz</orgName>
```

Ici, le fait d'insister sur les fonctions de M. de Marillac nous amènerait à privilégier la fonction au nom du personnage. Dans la première version du projet, cet encodage est en suspens.

En ce qui concerne les peuples comme les « Espagnols », les « Français », etc, nous constatons un vide à ce propos dans les Guidelines. Plus exactement, il est proposé de mettre l'élément <orgName> et l'attribut @type avec la valeur « peuple », ce qui ne nous semble pas vraiment satisfaisant. Notre décision est de ne pas les encoder.

D'autres problèmes se posent à nous :

- 1) les noms fictifs ou mythologiques
- 2) les noms de religion comme les congrégations, les églises, etc

Pour les premiers, nous avons choisi d'utiliser l'élément <name> et non <persName>. Par exemple, Saul est encodé ainsi :

```
<name>Saul</name>
```

Pour le second problème, il n'est pas simple de décider de l'encodage. Dans cette première phase d'encodage, nous avons remis la décision d'utiliser l'élément <rs> qui semble le plus adapté.

Enfin, nous reprenons l'élément <rs> avec son attribut @type pour qualifier les noms quand les

personnages ne sont pas cités expressément, par exemple : « sa fille » :

```
<rs type="person">sa fille</rs>
```

qui pourrait faire l'objet d'une régularisation :

```
<rs type="person">sa fille<reg>Creil, Marie de</reg></rs>
```

afin que ce personnage puisse être référencé dans l'index des noms de personnes.

Les lieux

Deux éléments ont été choisis pour marquer les noms de lieux : `<placeName>` et `<geogName>`. Nous avons délaissé `<address>` car les adresses sont toujours très approximatives dans nos sources du début du XVIIe siècle :

- `<placeName>` : quand il s'agit d'un lieu que l'on peut cartographier avec une latitude et une longitude, par exemple :

```
<placeName>Fontainebleau</placeName>
```

S'il s'agit d'une adresse approximative, nous utilisons également cet élément :

```
scize <placeName>rue Quiquempoix</placeName>
```

ou

```
<placeName>hostel de Nevers</placeName>
```

Les noms de pays et de régions sont encodés avec `<placeName>`, par exemple : France, Béarn.

- `<geogName>` : nous utilisons cet élément quand le nom est associé à une caractéristique géographique par exemple Île de Ré :

```
<geogName>Isle de Ré</geogName>
```

Pour les circonscriptions comme les évêchés et les généralités, nous avons décidé de ne pas les encoder en tant que tel dans la première phase du projet : cela nécessiterait la collaboration d'un spécialiste de l'administration sous l'Ancien régime.

5 L'indexation

Le projet Patin a la particularité de comporter une table élaboré par le scribe. Cette table peut être générée en TEI. Nous utilisons les éléments `<index>` et `<term>` (Patin, page 2) :


```
Mr. le Cardinal Bagni<index><term>Bagni</term></index>
```

L'index auto-généré correspond à la table du manuscrit original. Par défaut, l'encodage se fait en début de paragraphe dans lequel se trouve le mot indexé, suivant en cela la table originale. Cependant, parfois ce dernier n'a pas respecté sa règle ou a indexé deux fois le mot à deux pages différentes si le paragraphe se trouve coupé sur deux pages : dans ce cas, nous essayons de respecter la table et encodons là où se situe le mot indexé.

6 Les références croisés

Rappel :

En introduction, nous avons vu les différences entre les pointeurs, les liens, les ancrs, etc. Nous utilisons les éléments `<ref>` et `<ptr>` qui spécifient, dans le même texte, le renvoi d'une référence ou le fait de pointer vers une cible. L'attribut `@target` en précise la destination. Nous en avons d'ailleurs vu une exploitation possible avec l'élément `<quote>` pour signifier le locuteur, c'est-à-dire :

```
<quote type="letter">il devoit s'en servir, et que si la bienheureuse  
soeur Marie de l'Incarnation eust esté priée, et recommandée de chose  
semblable, elle y auroit obey<ptr target="#MDMA" type="author"/></quote>
```

Nous encodons alors l'expression ou le texte avec l'élément `<ptr>` et son attribut `@target` qui pointe sur l'identifiant de la personne indiqué dans le fichier externalisé (Marillac, chapitre 11, page 300).

Quant à l'élément `<ref>`, et comme nous l'avons vu avec les références bibliographiques, il est requis pour renvoyer la référence dans une liste bibliographique externalisée, ou encore vers un lien internet comme dans l'en-tête de la TEI :

```
<ref target="http://creativecommons.org/licenses/by-nc-nd/2.0/fr">  
Ce texte est protégé par la licence Creative Commons dont le contenu est  
disponible sous Creative Commons Attribution-Noncommercial-No Derivative  
Works 2.0 France  
</ref>
```

Index et renvois

Nous avons une seconde utilisation de ces éléments dans le projet Patin que nous détaillons ici :

- chaque paragraphe possède un identifiant que nous avons déclaré selon les critères suivants : les quatre premières lettres du mot principal indexé du paragraphe dans la table originale ; puis le numéro de la page :

```
<p xml:id="STRA_p3">
```

Ici, "STRA_p3" signifie « Strada » et « page 3 »

- les mêmes mots peuvent être indexés à d'autres endroits du texte ; si nous reprenons notre exemple, Strada est également indexé pages 70 et 75 avec des renvois page 70 sur les pages 3 et 75, comme suit :

```

Apres <persName><abbr>Mr.</abbr> de Thou</persName> le meilleur des
Historiens est ce me semble <persName>Famianus Strada</persName> Jesuite.
Son histoire est fort curieuse et fort réglée. Je voudrois bien qu'il
nous eust <note place="margin-left" type="note">Voyez plus haut
<abbr>pag.</abbr> <ref target="#STRA_p3">3.</ref> et plus bas
<abbr>pag.</abbr> <ref target="#STRA_p75">75.</ref></note> donné le 2.
Tome aussi beau que nous avons le premier

```

La référence de la page pointe vers la référence identifiée dans la table, identification qui correspond à celle du paragraphe. Le « # » indique que la référence est dans le même fichier.

Liens et références

Rappel : le meta-langage XML comporte la règle du non-chevauchement, inconvénient majeur :

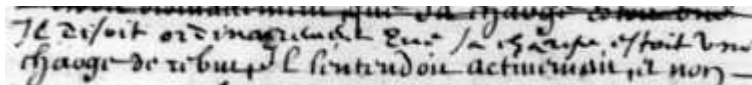
```

<add><del>Monsieur</del> <persName>M.</add> de Marillac</persName>

```

Cet encodage – fictif – n'est pas valide et ne permettra pas les traitements informatiques ultérieurs : nous l'avons donc barré.

Comment contourner le problème ? L'élément <join> est un élément possible, par exemple :



dont la transcription est la suivante :

```

Il disoit ordinairement, que sa charge estoit une charge de rebut (Marillac, chapitre 13, page 47v)

```

La première partie du discours, « sa charge estoit une », est un ajout qui oblige à fragmenter la phrase en deux « discours ». L'encodage suivant n'est pas valide :

```

<add>Il disoit ordinairement, que <q type="indirect" who="#MDMA">sa-
charge estoit une</add> charge de rebut.</q>

```

et doit être modifié ainsi :

```

<add>Il disoit ordinairement, que <q type="indirect" who="#MDMA">sa charge
estoit une</q></add> <q type="indirect" who="#MDMA3">charge de rebut.</q>

```

Pour l'édition finale de ce discours, il est nécessaire de joindre les deux parties avec l'élément <join> et les attributs @type, @target et @result, ce dernier permettant de produire le résultat du texte recomposé.

L'encodage initial ne change pas à l'exception de deux identifiants construits ainsi :

```
xml:id="MDMA-47v-1"
```

Ici, « MDMA » correspond à Michel de Marillac ; « 47v » à la page 47 verso ; « 1 » est le numéro d'ordre.

A l'encodage initial, on ajoute l'élément `<join>` avec ses attributs :

- `type="indirect"`
- `target="#MDMA-47v-1 #MDMA-47v-2"`
- `result="MDMA-47v"`

Ce nouvel attribut `@result` fait la somme des deux fragments :

```
<join type="indirect" target="#MDMA-47v-1 #MDMA-47v-2" result="MDMA-47v"/>
```

L'encodage final donne le résultat suivant :

```
<add>Il disoit ordinairement, que <q type="indirect" who="#MDMA"
xml:id="MDMA-47v-1">sa charge estoit une</q></add> <q type="indirect"
who="#MDMA" xml:id="MDMA-47v-2">charge de rebut.</q>
<join type="indirect" target="#MDMA-47v-1 #MDMA-47v-2" result="MDMA-47v"/>
```

Le but de cet encodage est de reconstituer l'intégralité des discours disséminés lors de la narration du texte par l'auteur.

La correspondance et l'alignement entre les textes

Dans le projet Marillac, trois versions du manuscrits sont encodées. Pour aligner ces versions et les faire correspondre, nous avons choisi une méthode, innovante, qui utilise les liens et les pointeurs entre les paragraphes des trois versions du manuscrit.

Cette méthode repose sur un processus qui fait appel à l'identification de chaque paragraphe pour constituer des liens entre paragraphe entre deux versions.

Identifier chaque paragraphe

Dans le texte, la structure de base choisie est le paragraphe ou le groupe de vers pour les poèmes. Un identifiant par paragraphe ou par groupe de vers doit être créé. Les identifiants sont composés de trois numéros : le premier est celui du chapitre précédé de la lettre "C" pour "chapitre" ; le second est celui de la page informatique trouvé dans l'élément `<pb>` dans chaque fichier, précédé des lettres "pb" ; le troisième et dernier numéro est celui du paragraphe de la page du manuscrit précédé de la lettre "P" pour "paragraphe". Voici deux exemples :

- dans le manuscrit BSG

```
<p n="11" xml:id="C1-pb8-P1">
```

- dans le manuscrit BN

```
<p n="11" xml:id="C1-pb9-P2">
```

Quand tous les identifiants sont créés et saisis dans les deux fichiers, nous pouvons procéder à la seconde phase.

Les liens

Nous avons créé un fichier externalisé de correspondance dans lequel nous utilisons les éléments `<linkGrp>` et `<link>` ainsi que l'attribut `@target` :

- `<linkGrp>` définit un ensemble de liens
- `<link>` est un lien
- `@target` est l'attribut qui pointe vers l'identifiant grâce à l'élément `<link>`

```
<link n="11" target="BSG_chap_11_mortifications.xml#C1-pb8-P1
BNF_chap_11_mortifications.xml#C1-pb9-P2"/>
```

La signification de ce lien est la suivante : « le paragraphe 1 de la page 8 du fichier numérique du manuscrit du chapitre 11 de la version du manuscrit de la Bibliothèque Sainte Geneviève correspond au paragraphe 2 de la page 8 du fichier électronique du manuscrit du chapitre 11 de la version du manuscrit de la Bibliothèque nationale de France ».

Afin de faciliter le travail et la compréhension, nous avons créé des entités.

Les entités dans XML

L'entité est une macro qui permet d'abrégé une expression, ici le nom du fichier. Par exemple, si nous gardons notre exemple simplifié, le nom des deux fichiers est simplifié ainsi :

```
<!ENTITY bsg11 "BSG_chap_11_mortifications.xml">
<!ENTITY bnf11 "BNF_chap_11_mortifications.xml">
```

Ces entités se placent dans le fichier XML/TEI juste après la formule déclarative :

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

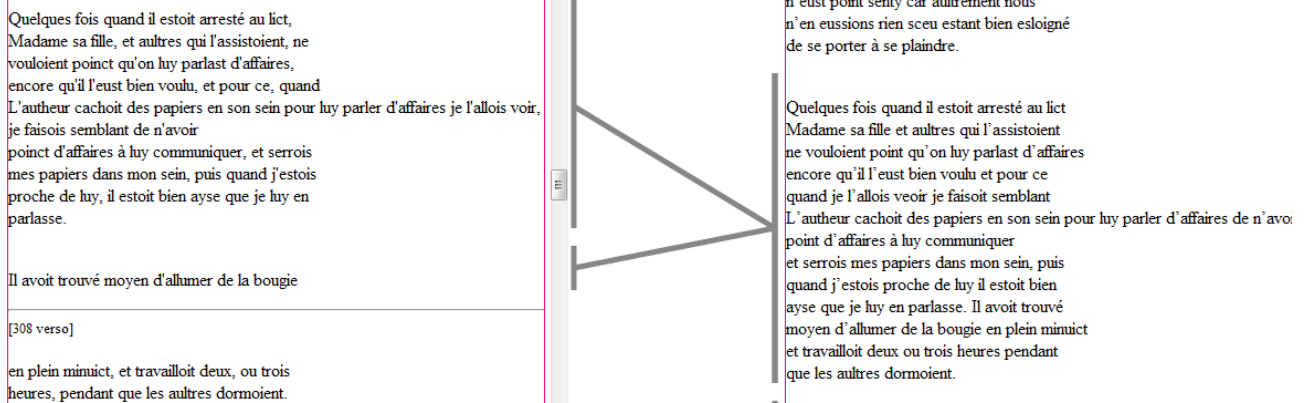
Les entités créés, il suffit de les réutiliser ainsi :

```
<link n="11" target="&bsg01;#C1-pb8-P1 &bnf01;#C1-pb9-P2"/>
```

Le résultat nous permet de faire correspondre les paragraphes :

1. pour montrer les variantes par paragraphe
2. pour « scroller » simultanément les deux versions.

Un premier résultat est donné ci-dessous :



Il est visible ici que le paragraphe qui commence ainsi « Quelque fois... » était scindé dans la seconde version BSG avant d'être réuni dans la troisième version BN.

Ce programme informatique gracieusement offert par Jean-Daniel Fekete, directeur de l'équipe Aviz à l'INRIA, est encore en cours d'achèvement.

L'analyse interprétative des textes

Dans une dernière phase d'analyse interprétative des textes du projet Marillac, nous utilisons les éléments `<interGrp>` et `<interp>` dans une grille interprétative externalisée. Dans le texte, nous devons alors délimiter une zone du texte avec un début et une fin : l'élément `<seg>` avec l'attribut `@ana` associé.

Cette partie du projet sera développée ultérieurement, dans la seconde phase du projet entre 2012 et 2014 par une équipe de chercheurs sous la responsabilité de Cécile Soudan, ingénieure de recherche dans le centre du Grihl²⁶, EHESS-CRH.

Conclusion

Le 5 mars 2012, nous avons organisé avec Cécile Soudan un colloque *Penser l'édition numérique savante. La TEI entre dévotion et libertinage : analyse de structure de textes du XVIIe siècle* à l'EHESS à Paris. Ce colloque présentait les premiers résultats des encodages des deux projets et particulièrement le démonstrateur TEI construit par Camille Harang, informaticien free-lance.

26 Voir le site du Grihl <http://www.ehess.fr/centres/grihl/>

Annexe 1 : fichier type du projet Marillac²⁷

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="projetsimple.xsl" type="text/xsl"?>
<!DOCTYPE TEI SYSTEM "../validation/tei_grihl_marillac.dtd">
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="fr">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          La Vie de Michel de Marillac par Nicolas Lefebvre de Lezeau - chapitre <!--
numéro --> : édition électronique format XML/TEI P5.
        </title>
        <editor xml:base="http://crh.ehess.fr">CRH-Grihl</editor><!-- éditeur
scientifique ? responsable du contenu intellectuel ? -->
        <principal xml:id="cs">Cécile Soudan</principal>
        <principal xml:id="vgl">Valérie Gratsac-Legendre</principal>
        <principal xml:id="nd">Nicole Dufournaud</principal>
        <funder xml:base="">
          Education, Audiovisual and Culture Executive Agency - EACEA - Education and
culture DG
        <date>2010-2014</date>
        </funder>
        <funder xml:base="http://www.ehess.fr"><name>Ecole des Hautes Etudes en
Sciences Sociales -
          EHESSE</name> Partenaire du projet : European Network for Baroque
          Cultural Heritage ENBaCH
        </funder>
        <funder xml:base=""><name>CNRS Délégation Paris A</name>
<address><addrLine>27 rue Paul Bert 94204 Ivry sur Seine
cedex</addrLine></address>
        </funder>
        <sponsor xml:base="">
          <name>Università Degli Studi di Roma "La Sapienza"</name> Coordinateur du
projet : European Network for Baroque
          Cultural Heritage ENBaCH
        </sponsor>
        <sponsor><name>Ecole des Hautes Etudes en Sciences Sociales -
          EHESSE</name> Partenaire du projet : European Network for Baroque
          Cultural Heritage ENBaCH
        </sponsor>
      </titleStmt>
      <editionStmt>
        <edition n="1">
          Edition électronique beta <date>2010</date> sur <ref
target="http://nicole.dufournaud.net/EHESSE/">le site personnel de Nicole
Dufournaud</ref>
        <date>2012</date>
        </edition>
      <respStmt>
        <resp>Chef de projet Enbach Paris</resp>
        <name>Cécile Soudan</name>
      </respStmt>
      <respStmt>
        <name>Nicole Dufournaud</name>
        <resp>Experte TEI - Analyste en humanités numériques</resp>
      </respStmt>

```

²⁷ Ce fichier type a été établi par Nicole Dufournaud.

```

    <respStmt>
      <name>Valérie Gratsac</name>
      <resp>Responsable de la transcription, analyste en humanités
numériques</resp>
    </respStmt>
    <respStmt>
      <name xml:id="fxp">François-Xavier Petit</name>
      <resp>Co-responsable de l'équipe scientifique et analyse</resp>
    </respStmt>
    <respStmt><!-- à modifier selon le chapitre analysé et mettre alors
Responsable de l'analyse -->
      <name xml:id="fm">Fabrice Micallef</name>
      <resp>Analyse</resp>
    </respStmt>
    <respStmt><!-- à modifier selon le chapitre analysé -->
      <name xml:id="cl">Camille Lefauconnier-Ripoll</name>
      <resp>Analyse</resp>
    </respStmt>
    <respStmt><!-- à modifier selon le chapitre analysé -->
      <name xml:id="ddl">Dora de Lima</name>
      <resp>Analyse</resp>
    </respStmt>
    <respStmt><!-- à modifier selon le chapitre analysé -->
      <name xml:id="AVS">Anne-Valérie Solignat</name>
      <resp>Analyse</resp>
    </respStmt>
  </editionStmt>
  <extent>
    <!-- taille du fichier en octets-->
  </extent>
  <publicationStmt>
    <!-- concerne la diffusion du texte électronique avec le lieu, les droits et
la date après le nom de l'autorité-->
    <publisher>
      <name xml:base="http://ehess.fr">Ecole des Hautes Etudes en Sciences
Sociales - EHESS</name>
      <address>
        <addrLine>190-198, avenue de France</addrLine>
        <addrLine>75013 Paris</addrLine>
      </address>
    </publisher>
    <authority>
      <name>François Weil</name>
      <affiliation>
        <name>EHESS</name>
      </affiliation>
    </authority>
    <pubPlace>Paris</pubPlace>
    <availability status="free" xml:base="http://ehess.fr">
      <p>
        Ce texte est protégé par la licence <ref
target="http://creativecommons.org/licenses/by-nc-nd/2.0/fr">Creative
Commons</ref> dont le contenu est disponible sous Creative Commons Attribution-
Noncommercial-No Derivative Works 2.0 France
      </p>
    </availability>
    <idno></idno> <!-- identifiant DOI ou URL à ajouter-->
    <date>2012</date>
  </publicationStmt>
  <sourceDesc>
    <!-- concerne la source originale -->

```

```

<msDesc><!-- Description de la source originale (manuscrit ou imprimé
ancien) -->
  <msIdentifier><!-- Identifie le manuscrit -->
    <country>France</country>
    <settlement>Paris</settlement>
    <institution><!-- AN ou BSG ou BNF --></institution>
    <repository><!-- par exemple Archives Privées --></repository>
    <collection/><!-- à compléter s'il y a lieu-->
    <idno type="cote"><!-- par exemple 1AP592 --></idno>
  </msIdentifier>
  <msContents><!-- décrit le contenu soit dans un <p> soit dans une structure
précise -->
    <summary><!-- par exemple Le manuscrit des Archives Nationales est le plus
ancien des cinq manuscrits connus encore conservés --></summary>
    <msItem><!-- lien vers le tableau des titres fait par CS sur le site
principal du projet -->
      <author>Nicolas Lefebvre de Lezeau</author>
      <title xml:base="">La vie de Michel de Marillac</title>
      <textLang mainLang="fre" otherLangs="lat"/>
    </msItem>
  </msContents>
  <physDesc><!-- décrit le contenant avec l'aide de Camille ?-->
    <objectDesc>
      <supportDesc><!-- Description matérielle des 3 éditions avec
dimension, taille, organisation des
feuilletts, l'état matériel -->
        <extent></extent><!-- compléter le nombre de feuilletts -->
        <foliation></foliation><!-- organisation des feuilletts -->
        <condition></condition><!-- l'état matériel -->
      </supportDesc>
    </objectDesc>
    <handDesc><!-- décrit les différentes écritures ainsi que les ajouts à
faire par manuscrit -->
      <p></p>
    </handDesc>
  </physDesc>
  <history>
    <!-- historique de la source voir Hélène sur l'historique de la source
faite pour AN. et BN-->
    <origin/>
    <provenance/> <!-- informe d'un épisode sur l'histoire du document entre
la création et l'acquisition -->
    <acquisition/>
  </history>
</msDesc>
</sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc xml:base=""><!-- résumé description du projet à faire -->
    <p></p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <!-- résumé : liste des corrections -->
      <p/>
    </correction>
    <hyphenation>
      <p>La coupure des lignes est respectée, à l'identique de celle du
manuscrit</p>
    </hyphenation>
    <normalization><!-- point sur la régularisation ou non du texte -->

```



```

    <p>Hormis les accents, les apostrophes, les majuscules, le texte n'a pas
fait l'objet de régularisation orthographique.</p>
</normalization>
<segmentation>
  <!-- déclaration sur la segmentation du texte -->
  <p>La ponctuation n'a pas fait l'objet d'un encodage particulier.</p>
</segmentation>
<interpretation>
  <p></p>
</interpretation>
</editorialDecl>
<classDecl>
  <taxonomy xml:id="bnf">
    <bibl>Répertoire d'autorité Rameau de la BNF</bibl>
  </taxonomy>
</classDecl>
</encodingDesc>
<profileDesc> <!-- description non-bibliographique -->
  <creation xml:base="">
    <!-- Histoire du projet électronique avec les 3 versions à faire par CS et
VGL-->
  </creation>
  <handNotes><!-- identification des scribes avec leur nom et leur rôle -->
    <handNote><!-- ex: Bignon ? --></handNote>
    <handNote></handNote>
  </handNotes>
  <langUsage><!-- Norme ISO 639-3 -->
    <language ident="fre" usage="80">Français</language><!-- signifie
le % de la langue française utilisé dans le fichier -->
    <language ident="lat">Latin</language>
    <language ident="gre">Grec</language>
    <language ident="gss">Grec ancien en alphabet grec</language>
  </langUsage>
  <textClass>
    <keywords scheme="#bnf">
      <list>
        <item>Religion</item>
        <item>Science politique</item>
        <item>Histoire</item>
      </list>
    </keywords>
  </textClass>
</profileDesc>
<revisionDesc>
  <change>
    <!-- A chaque étape importante de l'encodage du fichier, mettre la date, le
nom de l'encodeur et le type d'encodage formel, sémantique, etc -->
    <date> <!-- à compléter --></date>
    <name><!-- à compléter --></name>
    <!-- par exemple : encodage formel, encodage sémantique, analyse,
indexation -->
  </change>
</revisionDesc>
</teiHeader>
<text xml:lang="fre" type="chapter" n="numero">
  <!-- mettre dans @n le numéro du chapitre -->
  <pb/>
  <fw type="pageNum"></fw>
  <!-- S'il y a lieu, remplir le front -->
  <front>
    <titlePage>

```

```
<docTitle>
  <titlePart>
    <lb/>
  </titlePart>
</docTitle>
</titlePage>
</front>
<pb/>
<fw type="pageNum"></fw>
<pb/>
<fw type="pageNum"></fw>
<body>
  <div>
    <!--      Début du texte -->
    <head>
    </head>
  </div>
</body>
<back><!-- pièces annexes -->
  <div></div>
</back>
</text>
</TEI>
```

Annexe 2 : la feuille de style de travail (extrait)²⁸

```

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

<xsl:stylesheet
  xmlns:xd="http://www.pnp-software.com/XSLTdoc"
  xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0"
  xmlns:edate="http://exslt.org/dates-and-times"
  xmlns:estr="http://exslt.org/strings"
  xmlns:exsl="http://exslt.org/common"
  xmlns:fo="http://www.w3.org/1999/XSL/Format"
  xmlns:local="http://www.pantor.com/ns/local"
  xmlns:rng="http://relaxng.org/ns/structure/1.0"
  xmlns:tei="http://www.tei-c.org/ns/1.0"
  xmlns:teix="http://www.tei-c.org/ns/Examples"
  xmlns:html="http://www.w3.org/1999/xhtml"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:ntei="http://www.example.org/ns/nonTEI"
  extension-element-prefixes="exsl estr edate"
  exclude-result-prefixes="html xd exsl estr edate a fo local rng tei
  teix" version="1.0">

  <xd:doc type="stylesheet">
    <xd:short>
      Feuille de style pour traduire du "TEI/manuscrit" to HTML
    </xd:short>
    <xd:detail>
    </xd:detail>
    <xd:author>Jean-Daniel Fekete jean-daniel.fekete@inria.fr</xd:author>
    <xd:author>Nicole Dufournaud Nicole.Dufournaud@laposte.net</xd:author>
    <xd:cvsId>$Id$</xd:cvsId>
    <xd:copyright>2010, Ecole des Hautes Etudes en Sciences
    Sociales</xd:copyright>
  </xd:doc>

  <xsl:output method="html"
    encoding="UTF-8"
    doctype-public="-//W3C//DTD HTML 4.0 Transitional//EN"/>

  <xd:doc>
  <xd:short>Traduit l'élément head dans div</xd:short>
  <xd:detail>head prend le style</xd:detail>
  </xd:doc>
  <xsl:template match="tei:div/tei:head">
    <div>
      <xsl:call-template name="styling">
        <xsl:with-param name="style">font-size: 130%;</xsl:with-param>
      </xsl:call-template>
      <xsl:apply-templates/>
    </div>
  </xsl:template>

  <xd:doc>
  <xd:short>Traduit l'élément foreign</xd:short>
  <xd:detail>les langues étrangères sont dans une font italic</xd:detail>
  </xd:doc>

```

²⁸ Cette feuille de style a été créée par Jean-Daniel Fekete lors du projet Millefeuille dans le cadre d'une ACI entre les Archives nationales, l'Ecole nationale des chartes et l'université Paris I ; nous l'avons reprise en partie et l'avons modifiée. Nous en reprenons ici un extrait.

```

<xsl:template match="tei:foreign">
  <EM><xsl:apply-templates/></EM>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément p</xd:short>
</xd:doc>
  <xsl:template match="tei:p">
    <p>
      <xsl:call-template name="styling"/>
      <xsl:apply-templates/>
    </p>
  </xsl:template>

<xd:doc>
<xd:short>Traduit l'élément fw avec l'attribut place et la valeur top-
right</xd:short>
<xd:detail>le texte est à droite</xd:detail>
</xd:doc>
  <xsl:template match="tei:fw[@place='top-right']">
    <div>
      <xsl:call-template name="styling">
        <xsl:with-param name="style">position:relative ;text-align : right; font-
size: 80%;</xsl:with-param>
      </xsl:call-template>
      <xsl:apply-templates/>
    </div>
  </xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément lb</xd:short>
  <xd:detail>tei:lb transformé en html:br </xd:detail>
</xd:doc>
<xsl:template match="tei:lb">
  <br/>
  <xsl:apply-templates/>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément l</xd:short>
  <xd:detail>tei:l transformé en html:br</xd:detail>
</xd:doc>
<xsl:template match="tei:l">
  <br/>
  <span>
    <xsl:call-template name="styling"/>
    <xsl:apply-templates/>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément lg</xd:short>
  <xd:detail>tei:lg transformé en html:div</xd:detail>
</xd:doc>
  <xsl:template match="tei:lg">
    <p>
      <xsl:call-template name="styling"/>
      <xsl:apply-templates/>
    </p>
  </xsl:template>

```

```

<xd:doc>
  <xd:short>Traduit l'élément pb</xd:short>
  <xd:detail>tei:pb transformé en html:br et met une ligne de
    séparation pour montrer le saut de page</xd:detail>
</xd:doc>
<xsl:template match="tei:pb">
  <br/>
  <hr/>
</xsl:template>

<xd:doc>
<xd:short>Traduit l'élément space avec les attributs quantity et unit</xd:short>
</xd:doc>
  <xsl:template match="tei:space[@unit]">
    <SPAN style="width: '{@quantity}{@unit}'">
      <xsl:apply-templates/>
    </SPAN>
  </xsl:template>

  <xd:doc>
    <xd:short>Traduit l'élément note</xd:short>
    <xd:detail>La note est dans un rectangle sur la droite sans
frame.</xd:detail>
  </xd:doc>
  <xsl:template match="tei:note">
    <SPAN style="float: right; clear: right; width: 30%;">
      <xsl:apply-templates/>
    </SPAN>
  </xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément quote</xd:short>
  <xd:detail>La citation est en couleur verte.</xd:detail>
</xd:doc>
<xsl:template match="tei:quote">
  <span style="color:green; ">
    <xsl:if test="@xml:lang">
      <xsl:attribute name="style">font-style: italic;</xsl:attribute>
    </xsl:if>
    <xsl:apply-templates/>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément q (discours direct)</xd:short>
  <xd:detail>Le discours est encadrée par des guillemets.</xd:detail>
</xd:doc>
<xsl:template match="tei:q">
  <span>
    <xsl:if test="@xml:lang">
      <xsl:attribute name="style">font-style: italic;</xsl:attribute>
    </xsl:if>
    <xsl:text>&#171; </xsl:text>
    <xsl:apply-templates/>
    <xsl:text> &#187;</xsl:text>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément bibl</xd:short>
  <xd:detail>La référence bibliographique peu structurée dans une autre

```

```

langue que française est en italique.</xd:detail>
</xd:doc>
<xsl:template match="tei:bibl">
  <span>
    <xsl:if test="@xml:lang">
      <EM><xsl:attribute name="style"/></EM>
    </xsl:if>
    <xsl:apply-templates/>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément title dans bibl</xd:short>
  <xd:detail>Le titre d'une référence bibliographique est en italique et en
gras.</xd:detail>
</xd:doc>
<xsl:template match="tei:bibl/tei:title">
  <span>
    <xsl:call-template name="styling">
      <xsl:with-param name="style">font-weight: bold; </xsl:with-param>
    </xsl:call-template>
    <EM><xsl:apply-templates/></EM>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément expan</xd:short>
</xd:doc>
<xsl:template match="tei:expan">
  <span>
    <xsl:call-template name="styling">
      <xsl:with-param name="style">font-weight: bold; color:orange; </xsl:with-
param>
    </xsl:call-template>
    <xsl:apply-templates/>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément abbr</xd:short>
</xd:doc>
<xsl:template match="tei:abbr">
  <span>
    <xsl:call-template name="styling">
      <xsl:with-param name="style">font-weight: bold; color:orange; </xsl:with-
param>
    </xsl:call-template>
    <xsl:apply-templates/>
  </span>
</xsl:template>

<xd:doc>
  <xd:short>Traduit l'élément del</xd:short>
  <xd:detail>Les mots sont barrés simplement</xd:detail>
</xd:doc>
<xsl:template match="tei:del">
  <SPAN style="text-decoration:line-through; ">
    <xsl:apply-templates/>
  </SPAN>
</xsl:template>

```

```

<xd:doc>
<xd:short>Traduit l'élément sic</xd:short>
  <xd:detail>Le mot "sic" est ajouté entre crochets en petit après
    le mot marqué </xd:detail>
</xd:doc>
<xsl:template match="tei:sic">
  <SPAN>
  <xsl:apply-templates/>
  <SMALL><xsl:text>[sic]</xsl:text>
  </SMALL>
  </SPAN>
</xsl:template>

<xd:doc>
<xd:short>Traduit l'élément gap</xd:short>
<xd:detail>Le(s) mot(s) qui sont supprimés sont remplacés par trois
  points entre crochets </xd:detail>
</xd:doc>
<xsl:template match="tei:gap">
  <xsl:text>[...]</xsl:text>
</xsl:template>

<xd:doc>
<xd:short>Traduit l'élément add</xd:short>
  <xd:detail>Le(s) mot(s) ajouté(s) apparaissent sur la ligne en bleu
</xd:detail>
</xd:doc>
<xsl:template match="tei:add">
  <SPAN style="color: blue; ">
  <xsl:apply-templates/>
  </SPAN>
</xsl:template>

<xd:doc>
<xd:short>Traduit l'élément unclear</xd:short>
<xd:detail>Le mot "unclear" apparait entre crochets dans une petite
  font après le mot marqué</xd:detail>
</xd:doc>
<xsl:template match="tei:unclear">
  <SPAN style="color: red; ">
  <xsl:apply-templates/>
  <SMALL><xsl:text>[unclear]</xsl:text>
  </SMALL>
  </SPAN>
</xsl:template>

</xsl:stylesheet>

```

Annexe 3 : la grille d'analyse du projet Marillac (extrait)²⁹

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI SYSTEM "../validation/tei_grihl_marillac.dtd">
<TEI xml:lang="fr" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          La Vie de Michel de Marillac par Nicolas Lefebvre de Lezeau : édition
          électronique format XML/TEI P5.
        </title>
        <editor xml:base="http://crh.ehess.fr">CRH-Grihl</editor>
        <principal xml:id="cj">Christian Jouhaud</principal>
        <funder>Education, Audiovisual and Culture Executive Agency - EACEA -
          Education and culture DG</funder>
        <sponsor xml:base="">
          <name>Università Degli Studi di Roma "La Sapienza"</name> Coordinateur
          du projet : European Network for Baroque
          Cultural Heritage ENBaCH
        </sponsor>
        <sponsor>
          <name>Ecole des Hautes Etudes en Sciences Sociales -
          EHES</name> Partenaire du projet : European Network for Baroque
          Cultural Heritage ENBaCH
        </sponsor>
        </titleStmt>
        <editionStmt>
          <edition n="1">Edition beta
        <date>2011</date>
        </edition>
        <respStmt>
          <resp>Chef de projet Enbach Paris</resp>
          <name xml:id="cs">Cécile Soudan</name>
        </respStmt>
        <respStmt>
          <name xml:id="nd">Nicole Dufournaud</name>
          <resp>Analyste en humanités numériques</resp>
        </respStmt>
        <respStmt>
          <name xml:id="vg">Valérie Gratsac</name>
          <resp>Responsable de la transcription, analyste en encodage</resp>
        </respStmt>
        </editionStmt>
        <extent>
          <!-- taille du fichier en octets-->
        </extent>
        <publicationStmt>
          <!-- concerne la diffusion du texte électronique avec le lieu, les
          droits et la date après le nom de l'autorité-->
          <publisher>
            <name xml:base="http://ehess.fr">Ecole des Hautes Etudes en Sciences
            Sociales - EHES</name>
            <address>
              <addrLine>190-198, avenue de France</addrLine>
              <addrLine>75013 Paris</addrLine>
            </address>
          </publisher>
        </publicationStmt>
      </fileDesc>
    </teiHeader>
  </TEI>

```

²⁹ Cette grille d'analyse a été construite par Nicole Dufournaud. Depuis, elle a fait l'objet d'une révision importante par l'équipe de Cécile Soudan.


```

    </publisher>
    <authority>
      <name>Groupe de Recherches Interdisciplinaires sur l'Histoire du
Littéraire - GRIHL</name>
      <affiliation>
        <name>Centre de Recherches Historiques</name>
      </affiliation>
    </authority>
    <pubPlace>Paris</pubPlace>
    <availability status="free">
      <p>
        <ref target="http://creativecommons.org/licenses/by-nc-nd/2.0/fr">
          Ce texte est protégé par la licence Creative Commons dont le
contenu est disponible sous Creative Commons Attribution-Noncommercial-No
Derivative Works 2.0 France
        </ref>
      </p>
    </availability>
    <idno/>
    <!-- identifiant DOI ou URL à ajouter-->
    <date>2011</date>
  </publicationStmt>
  <sourceDesc>
    <p>Le manuscrit de "La Vie de Michel de Marillac par Nicolas
Lefebvre de Lezeau" a fait l'objet de plusieurs versions :
seules trois versions sont transcrites et encodées dans ce projet.</p>
  </sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc>
    <p>Une grille d'interprétation a été construite par une équipe
: François-Xavier Petit, Camille Lefauconnier, Dora de Lima,
Anne-Valérie Solognat, Fabrice Micallef, encadrée par Cécile
Soudan, Valérie Gratsac et Nicole Dufournaud.</p>
  </projectDesc>
</encodingDesc>
<revisionDesc>
  <change>
    <date>aout 2011</date>
    <name>Nicole Dufournaud</name>
    Création du fichier qui permet de dresser la liste des personnages
de l'entourage de Michel de Marillac
  </change>
</revisionDesc>
</teiHeader>
<text>
  <body>
<div>
<interpGrp resp="#FXP" type="corps">
<desc>Gestes, postures et déplacements du corps</desc>
<interp xml:id="I1A">Variations de l'espace et scénographie</interp>
<interp xml:id="I1B">Corps souffrant</interp>
<interp xml:id="I1C">Gestes et postures</interp>
<interp xml:id="I1D">Corps faible</interp>
<interp xml:id="I1E">Corps résistant</interp>
</interpGrp>
<gap/>
</div>
</body>
</text>
</TEI>

```

Annexe 4 : liste des « valeurs » d'attributs non variables³⁰

Typologie :

- Lettres ou missives : letter
- Arrêt : decision
- Eloge eloge
- Epitaphe : epitaphe
- Epigramme epigram
- Epitre epistle
- « Traité de la vie éternelle » traite
- Livre et manuscrit book
- Testament : testament
- Libelle libelle
- Poésie poetry
- Proverbe proverb
- Mémoire memoire
- Discours écrits speech
- Réclame catch
- Note simple note
- Entrée encyclopédique entry
- Numéro de page pageNum
- Discours direct direct
- Discours indirect indirect
- Auteur author
- Personnage person

Emplacement :

- Marge à droite : margin-right
- Marge à gauche : margin-left
- Liste ordonnée ordered

Rendu :

- Italic it

³⁰ Les valeurs des attributs @xml:id ou @spanTo par exemple ne sont pas donnés ici. Cette liste a été établie par les deux autrices et est susceptibles d'être modifiée lors de la seconde phase des projets.

- Trait pour supprimer un § `dash`
- Simple trait `strikethrough`
- Ronds `strikeCircle`
- Trait diagonal de gauche à droite `strikeDiagonalLR`
- Trait diagonal de droite à gauche `strikeDiagonalRL`
- Passages rayés par une croix `strikeCross`
- Trait vertical `strikeVertical`
- Double traits verticaux parallèle `strikeVerticaldouble`
- Double traits diagonal de gauche à droite
 `strikeDiagonalLRdouble`
- Double traits diagonal de droite à gauche
 `strikeDiagonalRLdouble`

Liste des éléments et attributs

| | |
|---------------------|--|
| Attributs..... | |
| @ana..... | 44 |
| @from..... | 34 sv |
| @ident..... | 26 |
| @n..... | 14 |
| @notAfter..... | 34 sv |
| @notBefore..... | 34 sv |
| @place..... | 11, 16 sv, 25 sv, 38 sv, 41, 51 |
| @quantity..... | 12, 15 |
| @reason..... | 15 |
| @rend..... | 21 sv, 25 |
| @rendition..... | 20 |
| @result..... | 41 sv |
| @seq..... | 24 sv |
| @spanTo..... | 8, 19 sv, 25 |
| @target..... | 8, 11, 20, 31 sv, 40 sv, 43 |
| @to..... | 34 sv |
| @type..... | 8, 11, 13 sv, 16, 19, 25 sv, 28 sv, 38, 41, 45, 50 |
| @unit..... | 11, 15 |
| @when..... | 34 sv, 38 |
| @where..... | 35 |
| @who..... | 8, 29 sv |
| +xml:id..... | 8, 19 sv, 25, 30, 35 sv, 40 |
| +xml:lang..... | 27, 32 |
| Eléments..... | |
| <abbr>..... | 10 |
| <add>..... | 11, 16 sv, 19 sv, 25, 38 |
| <address>..... | 39 |
| <addSpan>..... | 19 sv, 24 sv |
| <anchor>..... | 8, 19 sv, 25 |
| <author>..... | 11, 27, 32 |
| <back>..... | 10, 12 |
| <bibl>..... | 11, 27, 32 sv |
| <body>..... | 10, 13 |
| <cit>..... | 11, 30, 32 sv |
| <date>..... | 11, 34 sv, 38 |
| | 11, 16, 19 sv, 25 sv, 38 |
| <delSpan>..... | 21, 24 |
| <desc>..... | 11, 30, 35 sv |
| <div>..... | 10, 13 |
| <docTitle>..... | 12 |
| <event>..... | 11, 35 sv |
| <expan>..... | 10, 19 |
| <floatingText>..... | 10, 13, 31 |
| <foreign>..... | 11, 27, 32 |
| <front>..... | 10, 12 |
| <fw>..... | 10, 19, 25 sv |
| <gap>..... | 13, 20, 22 |
| <geogName>..... | 39 |

| | |
|-------------------|---------------------------------|
| <head>..... | 10 sv, 14, 35 sv |
| <index>..... | 39 |
| <interGrp>..... | 44 |
| <interp>..... | 44 |
| <item>..... | 14 sv |
| <join>..... | 8, 41 sv |
| <l>..... | 14, 31 |
| <label>..... | 11, 14, 30, 35 sv |
| <language>..... | 26 |
| <langUsage>..... | 26 |
| <lb>..... | 10, 12 sv, 16, 20 sv, 25, 31 sv |
| <lg>..... | 14, 31 |
| <link>..... | 8, 43 |
| <linkGrp>..... | 8, 43 |
| <list>..... | 14 sv |
| <listEvent>..... | 11, 35 sv |
| <listPerson>..... | 30 |
| <milestone>..... | 10 |
| <name>..... | 30, 38 |
| <note>..... | 10, 16, 41 |
| <orgName>..... | 38 |
| <p>..... | 10, 13 sv, 21 sv, 30, 40 |
| <particDesc>..... | 30 |
| <pb>..... | 10, 20, 25, 42 |
| <persName>..... | 11, 30, 37 sv, 41 |
| <person>..... | 30 |
| <personGrp>..... | 30 |
| <placeName>..... | 11, 38 sv |
| <ptr>..... | 8, 20, 31, 40 |
| <q>..... | 11, 28 sv |
| <quote>..... | 11, 13, 27, 30 sv, 40 |
| <ref>..... | 8, 32 sv, 40 sv |
| <roleName>..... | 37 sv |
| <rs>..... | 38 |
| <seg>..... | 8, 44 |
| <sic>..... | 11 |
| <space>..... | 11, 15 |
| <state>..... | 30 |
| <subst>..... | 25 |
| <supplied>..... | 15 |
| <teiHeader>..... | 7 sv, 10, 30 |
| <term>..... | 39 |
| <text>..... | 10, 12, 27 |
| <time>..... | 11, 34 sv |
| <title>..... | 11, 27, 32 sv |
| <titlePage>..... | 12 |
| <titlePart>..... | 12 |
| <trailer>..... | 12 sv |
| <unclear>..... | 10 |