Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits

Dernier enregistrement le 26 juin 2008

Alexei Lavrentiev (Alexei.Lavrentev@ens-lsh.fr)

CNRS / ENS-LSH, UMR 5191 ICAR

(VERSION 2.1- JUIN 2008)
(VERSION 2.0- MARS 2008)
(VERSION 1.0- DECEMBRE 2005)

Ce document de travail est élaboré dans le cadre des opérations de saisie et de traitement automatique des transcriptions de manuscrits de la Base de Français Médiéval (http://bfm.ens-lsh.fr). Le protocole du balisage est une extension du schéma de la TEI (P5) (http://www.tei-c.org) et se base sur les principes exposés dans le *Manuel d'encodage XML-TEI des textes de la Base de Français Médiéval* (http://bfm.ens-lsh.fr/IMG/pdf/Manuel_Encodage_TEI.pdf) et dans le *Manuel de description de textes pour la Base de Français Médiéval*, sur l'expérience du projet *Charrette* (http://www.princeton.edu/~lancelot/), sur les recommandations présentées dans le *Handbook* (v. 2) du projet Menota (http://www.menota.org/) et sur les propositions de l'initiative MUFI (http://gandalf.aksis.uib.no/mufi/).

Sommaire

Introduction : Principes du codage XML, application et extension de la TEI

Comme c'est le cas de l'ensemble des textes constituant la Base de Français Médiéval (BFM), le corpus de transcriptions de manuscrits, d'incunables et de livres imprimés du XVI^{e} siècle (BFM-MSS) est encodé selon la norme XML (http://www.w3c.org/xml), en conformité avec les recommandations de la Text Encoding Initiative (TEI, version P5, http://www.tei-c.org/Guidelines/P5).

Les principes de base du langage XML et de la TEI sont exposés dans la documentation de référence publiée sur les sites indiqués ci-dessus.

Les recommandations de la TEI étant très générales, une adaptation et une sous-spécification sont nécessaires dans le cadre d'un projet particulier. Le protocole d'encodage du corps des textes d'éditions modernes numérisées pour la BFM est exposé dans le *Manuel d'encodage XML-TEI des textes de la Base de Français Médiéval* (http://bfm.ens-lsh.fr/IMG/pdf/Manuel_Encodage_TEI.pdf). Le protocole de description et de caractérisation

des textes de la BFM est exposé dans le *Manuel de description de textes pour la Base de Français Médiéval*. Ces deux documents servent de base pour des extensions éventuelles dans le cadre d'un projet particulier.
La transcription fine de manuscrits et d'incunables a nécessité l'usage de balises spéciales, définies pour la plupart dans les modules **`transcr`** et **`msdescription`** de la TEI. Dans quelques cas précis, l'introduction de balises supplémentaires à celles proposées par la TEI a été jugée utile. Quelques unes de ces balises sont empruntées au schéma du projet Menota (Medieval Nordic Text Archive, http://www.menota.org), d'autres sont définies par le projet BFM-MSS. L'introduction de ces balises étrangères à la TEI s'est effectuée en conformité avec les mécanismes d'extension prévue par la TEI même.
L'espace de nommage des balises XML par défaut est celui de la TEI (http://www.tei-c.org/ns/1.0). Toutes les balises non-TEI appartiennent soit à l'espace Menota (http://www.menota.org/ns/1.0, préfixe **`me:`**), soit BFM (http://bfm.ens-lsh.fr/ns/1.0, préfixe **`bfm:`**).

1. Descripteur (entête TEI)

Le corpus est composé de transcriptions de fragments de manuscrits français médiévaux, d'incunables et de livres du XVI^{e} siècle. Chaque transcription forme une unité du corpus. Chaque unité du corpus porte un entête TEI.
Comme c'est le cas de l'ensemble des textes de la BFM, les descripteurs des unités textuelles sont stockés dans une base de données relationnelle, les entêtes TEI sont générés à partir des données de cette base au moment de la publication du corpus ou d'un document particulier. Par la publication nous entendons l'intégration à l'outil d'exploitation, toute mise en ligne, cession à un partenaire, etc.

Le descripteur d'une unité du corpus BFM-MSS inclut tous les champs caractérisant une œuvre dans la Base de Français Médiéval (titre, auteur, date de composition, forme, domaine, genre, etc.). La liste complète de ces champs est donnée dans le *Manuel de description de textes pour la BFM*.

La description du manuscrit et de sa transcription sont basées sur les dernières recommandations de la TEI, qui sont à leur tour, le fruit d'un long travail d'un groupe d'experts.
Un manuscrit médiéval (ou codex) peut en effet être composé de plusieurs parties hétérogènes réunies sous une reliure longtemps après l'écriture du texte. Chacune de ses parties peut à son tour inclure plusieurs unités textuelles (« œuvres » dans la terminologie de la BFM). Une transcription représente enfin une ou plusieurs parties d'une unité textuelle dans un manuscrit. Nous avons donc affaire à une structure hiérarchique, qui, dans une représentation XML prend forme d'arborescence d'éléments : **`<msDesc>`** (description du volume physique) peut contenir plusieurs **`<msPart>`** (parties physiquement homogènes) qui, à leur tour, contiennent un certain nombre de **`<msItem>`** (unités textuelles dans un manuscrit). Bien entendu, l'entête TEI d'une transcription donnée ne contiendra qu'un seul **`<msItem>`**.
Dans notre base de données relationnelle, cette hiérarchie est traduite par un système de tables entretenant une relation « un à plusieurs » et portant, pour la commodité, les mêmes noms que les balises TEI correspondantes (cf. le schéma des relations ci-joint). Enfin, une table spéciale ('MsTranscription') est dédiée aux données sur la transcription.

Les incunables et les livres du XVI^{e} siècle occupent une place intermédiaire entre les manuscrits et les éditions modernes : la description de l'exemplaire (objet physique localisé dans une bibliothèque) est pour eux aussi importante que les références de générale de

l'édition. Dans notre base de données ces imprimés anciens sont décrits dans des tables spécialisées ('Incunable', 'IncItem' et 'IncTranscription', cf. le schéma des relations). Dans l'entête TEI, les incunables sont décrits de la même façon que les manuscrits, avec quelques différences légères (par exemple, il n'y a pas de description des mains).

Un exemple d'un entête complet d'une transcription d'un manuscrit est présenté ci-dessous. Si la structure générale de l'entête est imposée par la TEI, les projets individuels disposent d'une certaine liberté dans le choix d'informations renseignées et dans leur disposition. Nous commenterons ensuite les éléments de l'entête dans l'ordre de leur apparition en indiquant nos choix aux cas où le schéma de la TEI est flexible.

```
<teiHeader type="text">
  <fileDesc>
    <titleStmt>
      <title type="normal">Chronique</title>
      <title type="source">Manuscrit Paris, BnF fr. 2682</title>
      <title type="reference" n="1012">monstre2682</title>
      <title type="medium">transcription électronique</title>
      <author>Enguerrand de Monstrelet</author>
      <editor role="editor">Équipe diachronie et bases textuelles d'ancien
et moyen français - UMR5191 ICAR, CNRS/ENS-LSH</editor>
      <respStmt>
        <resp>Transcription du manuscrit, encodage XML</resp>
        <name type="person" xml:id="AL">Alexei Lavrentiev</name>
      </respStmt>
      <respStmt>
        <resp>Datation du manuscrit</resp>
        <name type="document" xml:id="catBNF">Catalogue de la BnF</name>
      </respStmt>
    </titleStmt>
    <extent>- Taille approximative du fichier encodé en TEI non compressé :
104827 octets
- Nombre d'occurrences-mots : 847
- Numéros de pages saisies : de 233r à 233v
    </extent>
    <publicationStmt>
      <distributor><name>Projet BFM, UMR 5191 ICAR, CNRS/ENS LSH</name>
      <address>
        <addrLine>15 parvis René-Descartes</addrLine>
        <addrLine>B.P. 7000 </addrLine>
        <addrLine>69342 Lyon Cedex 07</addrLine>
        <addrLine>France</addrLine>
        <addrLine>Tél : 04 37 37 63 10</addrLine>
        <addrLine>Fax : 04 37 37 62 65</addrLine>
        <addrLine>http://bfm.ens-lsh.fr/</addrLine>
      </address>
      <email>bfm@ens-lsh.fr</email>
    </distributor>
    <availability status="restricted">
      <p>(c) 2007, Equipe BFM, CNRS/ENS-LSH.
        <hi>Conditions d'utilisation</hi> : usage interne équipe BFM
Pour d'autres conditions d'usage, contacter :
        <address>
          <addrLine><name type="person">Céline Guillot</name></addrLine>
          <addrLine>Mèl : Celine.Guillot@ens-lsh.fr</addrLine>
        </address>
      </p>
    </availability>
  </publicationStmt>
```

```
  <sourceDesc>
    <msDesc>
      <msIdentifier>
        <country>France;</country>
        <settlement>Paris</settlement>
        <repository>BnF</repository>
        <idno>fr. 2682</idno>
      </msIdentifier>
      <msPart>
          <altIdentifier>
            <idno>bnf_fr2682</idno>
          </altIdentifier>
          <physDesc>
            <handDesc hands="2">
              <handNote xml:id="scribe1">Premier scribe :
                <name type="person">inconnu</name>
              </handNote>
              <handNote xml:id="rubricateur1">Premier rubricateur :
                <name type="person">inconnu</name>
              </handNote>
            </handDesc>
          </physDesc>
          <history>
            <origin>
            Date du manuscrit : <origDate value="1450-01-01"
                                          notBefore="1400-01-01"
                                          notAfter="1499-01-01"
                                          resp="#catBNF"
                                          evidence="undefined"
                                          cert="low">15e s.
                                </origDate>
              Région : <region>inconnue</region>
              Ville : <settlement>inconnue</settlement>
              Atelier : <origPlace>inconnu</origPlace>
            </origin>
          </history>
        </msPart>
      </msDesc>
    </sourceDesc>
  </fileDesc>
  <profileDesc>
    <creation>
      <date when="1441-01-01">1441</date>
    </creation>
    <langUsage>
      <language ident="fr" usage="100">Le texte est entièrement écrit en
vers en français médiéval</language>
    </langUsage>
    <textDesc n="chronique">
      <channel mode="w"></channel>
      <constitution type="single"></constitution>
      <derivation type="original"></derivation>
      <domain type="historique"></domain>
      <factuality type="inapplicable"></factuality>
      <interaction type="none"></interaction>
      <preparedness type="prepared"></preparedness>
      <purpose type="inform"></purpose>
    </textDesc>
    <textClass>
      <catRef target="#forme_prose"/>
    </textClass>
```

```
  </profileDesc>
  <encodingDesc>
    <classDecl>
      <taxonomy xml:id="forme">
        <category xml:id="forme_vers">
          <catDesc>vers</catDesc>
        </category>
        <category xml:id="forme_prose">
          <catDesc>prose</catDesc>
        </category>
        <category xml:id="forme_mixte">
          <catDesc>mixte</catDesc>
        </category>
      </taxonomy>
    </classDecl>
    <projectDesc>
      <p>Projet : BFM-MSS - Base de Français Médiéval - Manuscrits
        Resp. : <name type="person">Alexei Lavrentiev</name>
        Équipe : Diachronie et bases textuelles d'ancien et de moyen
français
        Laboratoire : UMR5191 ICAR
        Institution : CNRS / ENS LSH, Lyon
      <address>
        <addrLine>15 parvis René-Descartes</addrLine>
        <addrLine>B.P. 7000 </addrLine>
        <addrLine>69342 Lyon Cedex 07</addrLine>
        <addrLine>France</addrLine>
        <addrLine>Tél : 04 37 37 63 10</addrLine>
        <addrLine>Fax : 04 37 37 62 65</addrLine>
        <addrLine>http://bfm.ens-lsh.fr/</addrLine>
      </address>
        <email>bfm@ens-lsh.fr</email>
      </p>
    </projectDesc>
    <editorialDecl>
      <p>See BFM-MSS Encoding Guidelines at http://bfm.ens-
lsh.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf.
See also BFM XML-TEI Encoding Guidelines at http://bfm.ens-
lsh.fr/IMG/pdf/Manuel_Encodage_TEI.pdf
and BFM XML-TEI Encoding Instructions at http://bfm.ens-
lsh.fr/IMG/pdf/Consignes_BFM.pdf</p>
    </editorialDecl>
  </encodingDesc>
  <revisionDesc>
    <change who="#AL" when="2007-07-22">Création de l'entête TEI et de la
représentation diffractée</change>
  </revisionDesc>
</teiHeader>
```

<teiHeader>

L'entête TEI peut contenir 4 éléments : **<fileDesc>**, **<encodingDesc>**, **<profileDesc>** et **<revisionDesc>**. Tous ces éléments sont présents dans l'entête BFM-MSS.

<fileDesc>

Description du document électronique. Dans le projet BFM-MSS cette description est composée de **<titleStmt>**, **<extent>**, **<publicationStmt>** et **<sourceDesc>**.

<titleStmt>

Déclaration du titre. Contient une ou plusieurs occurrences des éléments suivants : **`<title>`**, **`<author>`**, **`<editor>`**.

<title>

Balise générique pour un titre. Son attribut **`type`** est utilisé pour les catégoriser.

- **`type="normal"`** – titre conventionnel de l'œuvre, tel que donné dans les ouvrages de référence ;
- **`type="source"`** – brève indication de la source ;
- **`type="reference"`** – identifiant BFM (l'attribut **`n`** contient l'identifiant numérique dans le corpus de transcriptions) ;
- **`type="medium"`** – sert à préciser qu'il s'agit d'une transcription électronique.

<author>

Le nom de l'auteur.

<editor>

Balise générique pour un auteur « secondaire ». Son attribut **`role`** permet de préciser le rôle.

- **`role="editor"`** – institution responsable de l'édition du document numérique ;
- **`role="ed_sci"`** – éditeur scientifique (responsable des corrections dans la représentation normalisée) ;
- **`role="imprimeur"`** – scribe (utilisé dans **`<msDesc>`** pour les incunables).

Dans les deux derniers cas, cet élément porte également un attribut **`xml:id`** permettant de se référer à la personne correspondante.

<respStmt>

Déclaration des responsabilités, contient un ou plusieurs éléments **`<resp>`** et **`<name>`**.

<resp>

Description d'un rôle d'un responsable de l'édition.

<name>

Balise générique pour un nom propre. Son attribut **`type`** permet de les catégoriser. Son attribut **`xml:id`** permet de créer un identifiant auquel d'autres éléments pourront se référer (par exemple, d'indiquer le responsable d'une correction dans le texte).

<extent>

Informations sur les dimensions du document numérique.

<publicationStmt>

Déclaration de la publication : informations sur la distribution et l'accessibilité du document numérique. Contient les éléments **<distibutor>** et **<availability>** que nous ne détaillerons pas davantage.

<sourceDesc>

Conteneur de la description de la source du document numérique. Dans l'entête d'un texte du corpus BFM-MSS, il contient toujours un élément **<msDesc>**.

<msDesc>

Conteneur de la description du codex ou de l'incunable.
Il contient un **<msIdentifier>** et

- **<msPart>**, si le codex est hétérogène ;
- **<physDesc>** et **<history>**, si le manuscrit est homogène.

<msIdentifier>

Identifiant du manuscrit. Contient les éléments permettant de localiser le manuscrit pu l'exemplaire de l'incunable transcrit : **<country>** (pays), **<settlement>** (ville), **<repository>** (bibliothèque ou autre lieu de dépôt), **<idno>** (cote).

<msPart>

Contient un identifiant **<altIdentifier>** et les mêmes éléments que **<msDesc>** pour les manuscrits homogènes.

<physDesc>

Description physique du manuscrit ou de l'incunable. Contient l'élement **<objectDesc>** (description du volume en « prose libre ») et, pour les manuscrits, **<handDesc>**.

<handDesc>

Description des mains du manuscrit. Contient plusieurs éléments **<handNote>**. Son attribut **hands** sert à indiquer le nombre de mains. On distingue habituellement au moins deux mains : celles du scribe et du rubricateur (qui réalise les lettrines et les titres en couleur).

<handNote>

Une note caractérisant une main particulière du manuscrit. Son attribut **xml:id** permet de créer l'identifiant auquel on peut se référer pour indiquer la responsabilité de corrections scribales. La valeur de cet attribut indique le rôle et un indice numérique. Par exemple, **xml:id="scribe1"** pour le premier (ou unique) scribe identifié. Cet élément peut contenir de la « prose libre » avec un élément **<name>** qui indique éventuellement le nom de la personne.

<history>

Dans le corpus BFM-MSS, il contient un seul élément **<origin>**.

<origin>

Contient les informations sur l'origine du manuscrit ou de l'incunable :

- date du manuscrit (élément **<origDate>**, cf. élément **<date>** ci-dessous) ;
- région (élément **<region>**) ;
- ville (élement **<settlement>**) ;
- atelier pour les manuscrits (élément **<origPlace>**) ;
- nom de l'imprimeur pour les incunables (élément **<editor role="imprimeur">**).

<profileDesc>

Description du texte représenté par la transcription, indépendante de sa forme matérielle. Le même schéma est utilisé pour l'ensemble des textes de la BFM. Contient les éléments **<creation>**, **<langUsage>**, **<textDesc>** et **<textClass>**. Le modèle de cette description est commun pour l'ensemble des textes de la BFM.

<creation>

Informations sur la composition du texte, contient un élément **<date>** indiquant la date de composition.

<date>

Elément générique permettant d'encoder une date. D'autres éléments, plus spécialisés peuvent être utilisés selon les cas (par exemple, **<origDate>**), mais le mécanisme est toujours le même.
Le contenu de cet élément est l'indication de la date sous forme libre. Ses attributs permettent de formaliser la date pour la rendre exploitable par des outils informatiques :

- **when** : date au format ISO (aaaa-mm-jj) ;
- **notBefore** : la date la plus ancienne possible au format ISO ;
- **notAfter** : la date la plus tardive possible au format ISO ;
- **cert** : certitude de la datation ;
- **resp** : responsable de la datation.

L'algorithme du calcul de la date formalisée du projet BFM est décrit dans le *Manuel de description de textes pour la BFM*.

<langUsage>

Cet élément défini dans le module indique l'usage des différentes langues dans le texte. Contient un ou plusieurs éléments **<language>**.

<textDesc>

Caractérisation typologique du texte, contient les éléments **<channel>**, **<constitution>**, **<derivation>**, **<domain>**, **<factuality>**, **<interaction>**, **<preparedness>** et **<purpose>** dont l'usage est précidément défini par la TEI. Dans la BFM, le domaine du texte occupe une place centrale parmi les variables de la typologie textuelle. Une liste d'autorité des domaines a été établie et intégrée à la personnalisation du schéma de la TEI pour la BFM. Cette liste d'autorité figure dans le *Manuel de description de textes pour la BFM*.

D’autres variables importantes pour la BFM (telles que la forme et le genre du texte), elles ont donc été introduites à l’aide d’un mécanisme plus générique utilisant les éléments **<catRef>** et **<catDesc>** présentés ci-dessous.

<textClass>

Conteneur de la référence à la caractérisation du texte selon une taxonomie quelconque. Contient un ou plusieurs éléments **<catRef>**. Dans le projet BFM ce mécanisme est utilisé pour caractériser la forme (vers, prose ou mixte) et le genre du texte (roman, chronique, etc.). La liste d’autorité des genres et des formes figure dans le *Manuel de description de textes pour la BFM*.

<catRef>

Référence à une catégorie. Son attribut **target** assure la connexion avec la description de la catégorie **<catDesc>** dans le cadre d’une taxonomie.

<encodingDesc>

Description des normes d’encodage. Contient les déclarations des classes **<classDecl>**, ainsi que les éléments **<projectDesc>** et **<editorialDecl>** que nous ne détaillerons pas davantage.

<classDecl>

Déclaration d’une classe. Utilisée dans la BFM pour présenter les taxonomies de catégories de textes. Contient plusieurs éléments **<taxonomy>**.

<taxonomy>

Contient plusieurs éléments **<category>**.

<category>

Contient la description d’une catégorie **<catDesc>**. Son attribut **xml:id** permet de se référer à cette description à partir de **<catRef>**.

<catDesc>

Contient une description de la catégorie en « prose libre ».

<revisionDesc>

Historique des modifications du document numérique. Contient plusieurs éléments **<change>**.

<change>

Contient la description d’une opération de modification. Ses attributs **who** et **when** permettent d’indiquer la personne intervenante (qui doit figurer dans la déclaration des responsabilités) et la date de la modification.

2. Niveaux de représentation des données (diffraction)

Un document XML est organisé comme une arborescence d'éléments imbriqués. Ainsi, un roman en prose peut être représenté comme un élément **<text>** composé d'une partie liminaire (titre, épigraphe, prologue, etc.) et du « corps » du texte, composé à son tours de plusieurs chapitres. Les transcriptions des sources du projet BFM-Manuscrits visent à représenter de façon fine et fidèle les données des originaux, en y ajoutant des couches distinctes de normalisation et de « filtrage » permettant d'en faciliter la lecture et l'analyse. Pour atteindre cet objectif, il a été nécessaire de créer des représentations multiples des données et de choisir le niveau de la structure textuelle auquel ces représentations divergent.

Puisque la BFM est destinée avant tout à servir à des recherches linguistiques, l'unité principale de représentation qui s'impose est le mot. Le mot est en effet « porteur » des propriétés morpho-syntaxiques et l'unité d'indexation pour une grande partie des outils de l'informatique linguistique. Au même niveau que les mots se situent les marques de ponctuation et certaines balises de mise en page et de référence (sauts de ligne et de page, certains commentaires).
La structure de la transcription se complique dans les cas où le manuscrit présente des corrections scribales (suppressions ou ajouts) ou des passages illisibles (détérioration matérielle ou écriture ambigüe).

Nous utilisons le terme de *représentation diffractée*[1] pour désigner la séparation systématique des niveaux de transcription. C'est à partir de cette représentation que sont générées les différentes formes d'éditions (visualisations) et c'est à cette représentation que sont rattachées les annotations linguistiques. En revanche, une *syntaxe compacte* est utilisée au stade de saisie et de correction des transcriptions (cf. section 5 ci-dessous).

Illustration de la structure arborescente

```
<text>
      <front>
      </front>
      <body>
            <div>
                  <p>
                        <w><choice>
                              <me:norm>chevaliers</me:norm>
                              <me:dipl>ch<ex>eualie</ex>rs</me:dipl>
                              <me:facs>
                                <bfm:mdvAbbr>ch&apos;&rrot;</bfm:mdvAbbr>s
                              </me:facs>
                              <me:pal> … </me:pal>*[2]
                        </choice></w>
                        <bfm:punct><choice>
                              <me:norm></me:norm>
                              <me:dipl>,</me:dipl>
                              <me:facs>&punctelev;</me:facs>
                              <me:pal> … </me:pal>*
                        </choice></bfm:punct>
                  </p>
            </div>
      </body>
```

[1] Terme proposé par Nicolas Mazziotta.
[2] Dans les exemples de balisage, un astérisque après une balise signifie que l'usage de cette balise est optionnel.

```
</text>
```

Représentation normalisée : <me:norm>

Ce niveau de représentation est conforme aux normes appliquées dans les éditions critiques modernes (cf. Vielliard, Guyotjeannin 2001). La seule différence est qu'il s'agit toujours de la transcription d'un manuscrit, par conséquent toutes les corrections basées sur des leçons des autres manuscrits ou sur des données extratextuelles (corrections de noms géographiques, personnages historiques, etc.) ne sont pas prises en compte. Celles-ci peuvent éventuellement figurer dans des commentaires (élément **<note>**).

Traits caractéristiques

- Usage de caractères modernes (suppression des variantes calligraphiques et des diacritiques médiévaux) ;
- Résolution tacite des abréviations ;
- Introduction de la distinction *u/v, i/j* ;
- Usage de diacritiques modernes désambigüisants (*é, ï, ü,* etc.) ;
- Séparation de ligatures ;
- Notation des noms propres avec les majuscules ;
- Segmentation des unités-mots normalisée ;
- Ponctuation modernisée.

Contenu :

text (PCDATA)
<corr>
<supplied>

Représentation diplomatique : <me:dipl>

Cette représentation se situe dans la lignée des éditions dites « diplomatiques », même si les pratiques varient beaucoup selon les éditeurs. Les normes appliquées dans le projet BFM-MSS se basent sur le principe de transcription graphématique : les variantes de graphèmes sont neutralisées, mais on n'introduit pas de marques désambigüisantes.

Traits caractéristiques

- Usage de caractères modernes (suppression des variantes calligraphiques et des diacritiques médiévaux) ;
- Résolution des abréviations avec une mise en évidence typographique des caractères restitués ;
- Les caractères *i, j, u* et *v* du manuscrit sont respectés ;
- Pas d'usage de diacritiques modernes désambigüisants (*é, ï, ü,* etc.) ;
- Séparation de ligatures ;
- Respect des majuscules et des minuscules du manuscrit ;
- Segmentation des unités-mots normalisée ;
- Représentation de la ponctuation de la source : une virgule pour une ponctuation faible et un point pour une ponctuation forte.

Contenu :

texte (PCDATA)
<bfm:expan>
<ex>
<sic>

Représentation imitative : <me:facs>

Cette représentation s'inspire de certaines éditions dites « hyper-diplomatiques » ou « imitatives ». Son principe est de représenter fidèlement toutes les données potentiellement pertinentes pour une analyse linguistique du système graphique. Cette représentation ne peut remplacer ni la représentation normalisée, ni la représentation photographique de la source, mais elles les complète en répondant à des besoins de recherches spécifiques.

Traits caractéristiques

- Distinction de variantes calligraphiques clairement identifiables (allographes), maintien de diacritiques médiévaux ;
- Représentation des marques d'abréviation médiévales ;
- Les caractères *i, j, u* et *v* du manuscrit sont respectés ;
- Pas d'usage de diacritiques modernes désambigüisants (*é, ï, ü,* etc.) ;
- Maintien de certaines ligatures ;
- Représentation des majuscules et des minuscules du manuscrit (éventuellement, des « grandes minuscules » et des « petites majuscules ») ;
- Segmentation des unités-mots originale (maintien des agglutinations et des déglutinations) ;
- Marques de ponctuation médiévales ;
- Représentation des marques de correction scribales.

Contenu (mots) :

texte (PCDATA)
<add>
<am>
<bfm:mdvAbbr>
<bfm:lettrine>
<del>
<unclear>

Contenu (ponctuation)

texte (PCDATA)
<bfm:eol>

Représentation paléographique (facultative) : <me:pal>

Niveau de représentation prévu pour les annotations paléographiques (distinction fine des tracés de caractères, des traits d'écriture individuels, des espaces de différentes tailles, etc.). Il n'est pas implémenté à présent dans les transcriptions BFM-MSS.

3. Corrections scribales et éditoriales ; passages difficilement lisibles ou illisibles

Les transcriptions du projet BFM-Manuscrits visent à représenter les manuscrits sans corriger les erreurs éventuelles des scribes ou copistes. Par conséquent, les seules corrections y intégrées de façon systématique sont celles effectuées sur le manuscrit par le scribe lui-même ou par un (re)lecteur médiéval. Des corrections éditoriales de coquilles évidentes peuvent toutefois être ajoutées au niveau de représentation normalisée. Les mêmes principes de balisage s'appliquent aux passages du manuscrit difficilement lisible ou illisible à cause de l'état matériel du manuscrit. Les balises XML recommandées par la TEI sont utilisées pour toutes les corrections et passages illisibles.
Le niveau hiérarchique sur lequel apparaissent les balises de corrections et assimilées dépend de l'étendu de celles-ci. Trois cas de figure se présentent :

Partie d'un mot

Les éléments correspondants (**<corr>**, **<add>**, **<del>**, **<unclear>**, etc.) sont placés à l'intérieur de la représentation facsimilaire, diplomatique ou normalisée selon les cas.

Exemple

```
<w><choice>
      <me:norm>Platon</me:norm>
      <me:dipl>platon</me:norm>
      <me:facs>p<del rend="dots_below"
hand="#scribe1">i</del>laton</me:facs>
</choice></w>
```

Un mot entier ou plusieurs mots

Les mots (**<w>**) concernés sont placés dans les éléments correspondants (**<corr>**, **<add>**, **<del>**, **<unclear>**, etc.). Dans les passages supprimés, seul le niveau de représentation facsimilé sera présent.

Exemple

```
<del rend="dots_below" hand="#scribe1">
      <w type="num">
            <me:facs>.iij.</me:facs>
      </w>
      <w>
         <bfm:mdvAbbr><me:facs>phi<am>&apos;.</am><bfm:mdvAbbr></me:facs>
      </w>
      <space dim="horizontal" unit="chars" quantity="9"/>
      <w>
            <me:facs>in</me:facs>
      </w>
      <w>
            <me:facs>p&rrot;ologo</me:facs>
      </w>
</del>
```

Segment commençant ou finissant au milieu d'un mot et s'étendant sur plusieurs mots

Ce cas, assez rare, présente un problème de structure hiérarchique XML. La technique de « fragmentation et reconstitution d'éléments virtuels » est utilisée pour le résoudre (cf. *TEI Guidelines*, chapitre 20, section 3), qui semble satisfaisante pour le traitement de cas rares.

Exemple

```
<w>
      <norm>lairons</me:norm>
      <me:dipl>l<unclear xml:id="ucl123"
next="#ucl124">airons</unclear></me:dipl>
      <me:facs>l<unclear corresp="#ucl123" next="#ucl124">airons</unclear>
      </me:facs>
</w>
<unclear xml:id="ucl124" prev="#ucl123">
   <w>
      <norm>nous</me:norm>
      <me:dipl>nous</me:dipl>
      <me:facs>nous</me:facs>
   </w>
</unclear>
```

Notes et commentaires du relecteur/encodeur

Une **<note>** peut être utilisée pour ajouter un commentaire à l'endroit du texte où l'encodage pose un problème ou si le choix de balisage nécessite une explication. L'élément **<note>** est normalement placé au même niveau d'arborescence XML que les mots (**<w>**), à la fin du passage commenté.

4. Définition et description des éléments

Par défaut, les éléments ci-dessous sont définis par la TEI. Les éléments définis par le projet Menota portent un préfixe **me:** (**xmlns:me="http://www.menota.org/ns/1.0"**). Les éléments définis par le projet BFM-MSS portent le préfixe **bfm:** (**xmlns:bfm="http://bfm.ens-lsh.fr/ns/1.0"**).

<ab>

Balise générique pour une unité textuelle de niveau équivalent au paragraphe. Dans la BFM, cette balise est utilisée dans les textes en vers.

> La structure linguistique (en particulier, la division en phrases ou « unités ponctuables » est prioritaire dans la BFM. Pour éviter le conflit de « hiérarchies concurrentes », les éléments **<l>** (vers) et **<lg>** (groupe de vers) ne sont donc pas utilisés.

Position (« parents » ou « frères et sœurs »)

A l'intérieur d'une division (** ** ou autre).

Contenu (« enfants »)

<w>
<bfm:punct>
<lb>, etc.

Attributs

- **type** : type d'unité (**"gv"** pour groupe de vers) ;
- **n** : indique éventuellement le numéro du paragraphe.

<add>

Lettres, mots ou groupes de mots ajoutés au texte initial dans le manuscrit par le scribe ou autre personne à l'époque proche de la création du manuscrit.

Attributs

- **place** : indique l'endroit où le texte ajouté est placé. La TEI propose une liste de valeurs non restrictive (**"infralinear"**, **"supralinear"**, **"margin-right"**, **"inline"**, etc.). Une valeur **"overwritten"** est ajoutée par le projet BFM-MSS. Attribut obligatoire ;
- **hand** : indique la main ayant effectué l'ajout. Si absent, on suppose que c'est le premier scribe ;
- **resp** : indique le responsable du balisage et d'attribution de l'ajout
- **seq** : indique éventuellement l'ordre séquentiel en cas de plusieurs « couches » de corrections dans un manuscrit.

Position et contenu

Cf. section 3 ci-dessus.

<am>

Marque d'abréviation

Position

Dans l'élément **<bfm:mdvAbbr>** (représentation facsimilaire).

Contenu

PCDATA (un caractère MUFI représenté par une entité, cf. section 6).

Attributs

- **resp** – responsable du balisage (identification de la marque).

Exemples

Cf. **<bfm:mdvAbbr>** plus bas.

<bfm:headlb>

Saut de ligne « supplémentaire » dans un titre.

Cet élément est utilisé dans un cas particulier de la mise en page médiévale. Dans certains manuscrits les titres ou résumés de chapitres sont écrits en encre de couleur différente et placés immédiatement après la fin du chapitre précédent, souvent sans retour à la ligne. Parfois ces titres continuent sur les fins des premières lignes du chapitre suivant, ce qui pose un problème de continuité linéaire du texte (cf. illustration ci-dessous). La solution proposée est de placer la totalité du titre avant le début du chapitre annoncé et d'utiliser une balise spéciale la où une partie du titre se trouve sur la même ligne physique que le contenu d'une autre unité sémantique du texte (fin de la division précédente ou début de la division suivante). En revanche une simple balise **<lb>** est utilisée lorsqu'une ligne du titre occupe la totalité d'une ligne physique.

Fragment de l'*Image du monde*, ms. Paris, BNF, fr. 574, f° 83r (cf. exemple de code ci-dessous)

Position

Dans un titre (ou résumé) d'un chapitre ou autre unité textuelle (**<head>**).

Comme **<lb>**, cet élément marque en effet le début (et non la fin d'une ligne). La première balise se trouve donc au début du titre.

Contenu

Elément vide.

Attributs

- **lines** – nombre de lignes de décalage par rapport à la position normale (fin de la dernière ligne de la division précédente. Sa valeur est **"0"** si le titre commence à la même ligne, elle est négative si le titre commence avant la dernière ligne de la division précédente et positive si le titre continue sur des fins de ligne du début de la division suivante.

Exemple

(cf. illustration ci-dessus)

```
<lb/>est desus ia ne sera si fo<bfm:lb/>rt
<lb/>tour . </p></div>
<div><head><bfm:headlb lines="0"/>comment la terre
<bfm:headlb lines="1"/>crolle et fent . </head>
<p>
<lb/>ORe
<lb/>entendez donques
<lb/>du mouuement que
```

<bfm:hyphen>

Marque de coupure de mot en fin de ligne

Position dans la structure

Représentation facsimilaire : **`<me:facs>`**.

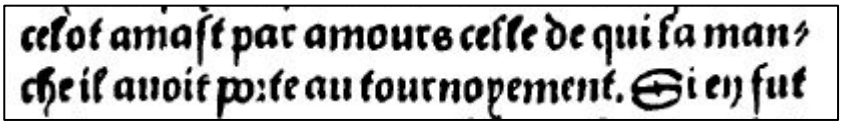
Contenu

Un caractère (représenté souvent par une entité MUFI **`&sol;`**, **`&dbloblhyphen;`**)

Attributs

- **`place`** – indique éventuellement la position de la marque de coupure (**`"margin"`** ou, par défaut, **`"inline"`**).

Exemples



```
la man<bfm:hyphen>&dbloblhyphen;</bfm:hyphen><lb/>che
```

<bfm:lettrine>

Lettre initiale (lettrine).

Position dans la structure

Représentation imitative : **`<me:facs>`**.

Contenu

Un caractère (PCDATA)

Attributs

- **`color`** – couleur (principale) de l'initiale ;
- **`colorSuppl`** – couleur secondaire ;
- **`size`** – taille prévue : nombre de lignes d'alinéa ;
- **`sizeAct`** – taille réelle de l'initiale en nombre de lignes ;
- **`decoration`** – détails de la décoration ;
- **`ref`** – référence de la description complète de l'initiale ;
- **`resp`** – responsable du balisage.

Exemples

```
<bfm:lettrine color="blue" colorSuppl="red" size="6"
sizeAct="6">O</bfm:lettrine>R
```

<bfm:mdvAbbr>

Abréviation médiévale. Cette balise marque le segment graphique « affecté » par l'abréviation. Selon le type d'abréviation, il peut s'étendre de la marque d'abréviation proprement dite à une ou plusieurs lettres associées à la marque, voire à un mot entier. Le fait que les limites de l'abréviation ne coïncident pas forcément avec celles d'un mot nous empêche d'utiliser l'élément **<abbr>** de la TEI.

Position dans la structure

Dans la représentation facsimilaire : **<me:facs>**

Contenu

caractères (PCDATA)
<am>

Attributs

- **range** – précision de l'étendu de l'abréviation (**"chars"**, **"word"** or **"phrase"**) ;
- **type** – type d'abréviation (**"regular"**, **"frequent_word"**, etc.) ;
- **resp** – responsable du balisage (en particulier, de la délimitation de l'abréviation) ;
- **cert** – certitude de l'identification de l'abréviation.

Exemples[3]

```
<bfm:mdvAbbr><am>&et;</am></bfm:mdvAbbr>
h<bfm:mdvAabbr>o<am>&bar;</am></bfm:mdvAbbr>me
<bfm:mdvAbbr>ml<am>&apos;</am>t</bfm:mdvAbbr>
```

[3] Tous les exemples cités sont tirés du corpus BFM-MSS. Pour améliorer la lisibilité, seules les balises pertinentes pour l'exemple en question sont conservées. Sauf indication contraire, les graphies sont représentées au niveau diplomatique.

<bfm:punct>

Marque de ponctuation ou toute « frontière ponctuable ».

Cette balise sert à la fois à représenter les marques de ponctuation utilisées dans le manuscrit, les marques ajoutées au niveau « normalisé » pour faciliter la lecture du texte et à repérer les « frontières ponctuables ».

Une balise **<me:punct>** est proposée par le projet Menota, mais elle n'a pas exactement les mêmes propriétés que la balise **<bfm:punct>**.

Position

Au même niveau que les mots (**<w>**).

Contenu

Un élément **<choice>** ou l'une des représentations multiples : **<me:norm>**, **<me:dipl>**, **<me:facs>**.

Attributs

- **force** : indique la force de la ponctuation ("strong" ou "weak") ;
- **mark** : donne le nom conventionnel de la marque de ponctuation médiévale, sa valeur est "none" si aucune marque de ponctuation médiévale n'est présente dans la source (une marque peut apparaitre en revanche au niveau normalisé ou diplomatique) ;
- **ana** : le code du type de frontière ponctuable.

<bfm:sb>

Espace blanc à l'intérieur d'un mot (« déglutination »).

Position dans la structure

Dans la représentation facsimilaire : **<me:facs>** d'un mot.

Contenu

Elément vide

Attributs

- **cert** : certitude de la présence d'un espace ;
  NB : L'absence de certitude de déglutination signifie normalement qu'il y a un « petit blanc ».
- **resp** : responsable du balisage.

Exemples

```
ce quil a<mdv_sb cert="no"/>fiert a
```

<cb>

Saut de colonne.

Position et contenu

Elément vide, situé entre des mots ou des divisions de niveaux supérieurs.

Attributs

- **`n`** : indique le numéro de la colonne (habituellement **`"a"`**, **`"b"`**, **`"c"`** ou **`"d"`**).

<choice>

Balise TEI servant à délimiter les représentations alternatives d'un même segment du texte. Dans le projet BFM-MSS, à l'exemple du projet Menota, cette balise entoure les représentations alternatives d'un mot ou d'une marque de ponctuation.

Position

A l'intérieur d'un **`<w>`** ou d'un **`<bfm:punct>`**.

Contenu

`<me:norm>`, **`<me:dipl>`** et **`<me:facs>`**.

Attributs

Pas d'attributs.

<corr>

Lettres, mots ou groupes de mots corrigés par l'éditeur ou par un relecteur/encodeur dans la représentation normalisée.

Position et contenu

Cf. section 3 ci-dessus.

Attributs

- **`cert`** : indique la certitude de la correction ;
- **`resp`** : indique le responsable du balisage et d'attribution de la correction.

<damage>

Fragment du manuscrit endommagé.

Position et contenu

Cf. section 3 ci-dessus.

Attributs

- **`extent`** : indication de la taille du fragment endommagé ;
- **`resp`** : indique le responsable du balisage et d'attribution de la correction.

<del>

Lettres, mots ou groupes de mots supprimés dans le manuscrit.

Position et contenu

Cf. section 3 ci-dessus.

Attributs

- **`rend`** : indique la méthode de suppression utilisée. Les valeurs utilisées sont les suivantes : **`"dots_below"`**, **`"horbar"`**, **`"diagbar"`**, etc. ;
- **`hand`** : indique la main ayant effectué l'ajout, si identifiable ;
- **`resp`** : indique le responsable du balisage et d'attribution de l'ajout.

* *

Division structurelle du texte.

Position

A l'intérieur du corps du texte **`<body>`** ou d'une autre **`<div>`**.

Contenu

Eventuellement **`<head>`** au début ;
`<p>` ou **`<ab>`**.

Attributs

- **`rend`** : indique la méthode de suppression utilisée. Les valeurs utilisées sont les suivantes : **`"dots_below"`**, **`"horbar"`**, **`"diagbar"`**, etc. ;
- **`hand`** : indique la main ayant effectué l'ajout, si identifiable ;
- **`resp`** : indique le responsable du balisage et d'attribution de l'ajout.

<ex>

Lettres restituées dans la résolution d'une abréviation.

Position dans la structure

Dans la représentation diplomatique.

Contenu

Une ou plusieurs lettres (PCDATA)

Attributs

- **`cert`** : certitude de la résolution ;
- **`resp`** : responsable de la restitution des lettres.

Exemples

```
ch<ex>eualie</ex>r
ih<ex>e</ex>r<ex>usa</ex>l<ex>e</ex>m
```

<gap>

Fragment de texte omis : soit parce qu'il est illisible dans le manuscrit, soit pour des raisons d'échantillonnage.

Position et contenu

Cf. section 3 ci-dessus.

Attributs

- **`unit`** : indique l'unité de mesure : **`"chars"`**, **`"words"`**, **`"pages"`**, etc. ;
- **`extent`** : indication numérique de l'étendu (en unités choisies) ;
- **`reason`** : indique la raison de l'omission : **`"illegible"`**, **`"deleted"`** ou **`"sampling"`** ;
- **`hand`** : en cas d'effacement volontaire, indique le responsable, si identifiable ;
- **`resp`** : indique le responsable du balisage et d'attribution de l'ajout.

<head>

Titre ou autre élément liminaire d'une division textuelle (chapitre, partie, etc.).

Position et contenu

Début d'une division **`<div>`**.

Contenu

Mots **`<w>`** et autres unités du même niveau ;
`<bfm:headlb>` dans un cas particulier (cf. la définition de cet élément).

Attributs

Attribut globaux seulement.

<hi>

Segment de texte mise en relief.

Position et contenu

Cf. section 3 ci-dessus.

Attributs

- **`rend`** : indique la forme de mise en relief : **`"exp"`**, **`"color:red"`**, etc.

<lb>

Saut de ligne.

Position et contenu

Elément vide, situé entre des mots (sauf car particulier).

> Contrairement à l'apparence, cette balise marque un début et non la fin d'une ligne. Elle est par conséquent placée au début de la première ligne et est absente à la fin de la dernière.

Attributs

- **n** : indique le numéro de la ligne dans la page ou dans la colonne.

Cas particulier : saut de ligne coupant un mot.

L'attribut **ed="facs"** est utilisé à l'intérieur de la représentation imitative à l'endroit précis du saut de ligne dans le manuscrit. Une deuxième balise **<lb>** est placée après le mot coupé, portant un attribut **ed="norm"**.

Exemples

```
<lb n="4"/> que parfaitte proëce estoit entee et en<lb ed="facs"/>rachinee
<lb ed="norm" n="5"/> au plus fort ez cuers des nobles
```

<me:dipl>

Représentation diplomatique d'un mot ou d'une marque de ponctuation.

Position et contenu

Cf. section 0 ci-dessus.

<me:facs>

Représentation « facsimilé » d'un mot ou d'une marque de ponctuation. Elément défini par le projet Menota.

Position et contenu

Cf. section 0 ci-dessus.

<me:norm>

Représentation normalisée d'un mot ou d'une marque de ponctuation.

Position et contenu

Cf. section 0 ci-dessus.

<me:pal>

Représentation « paléographique » d'un mot ou d'une marque de ponctuation. Elément défini par le projet Menota.

Le projet BFM-MSS ne prend pas en charge ce niveau de représentation au stade actuel.

<note>

Note ou commentaire du relecteur/encodeur. Elément défini par la TEI.

Position dans la structure

A l'intérieur de paragraphes **<p>**.

Contenu

Texte brut (métatexte par rapport à la transcription).

<p>

Paragraphe (dans les textes en prose).

Position

A l'intérieur d'une division (** ** ou autre).

Contenu

<w>
<bfm:punct>
<lb>, etc.

Attributs

- **n** : indique éventuellement le numéro du paragraphe.

<pb>

Saut de page (ou de folio).

Position

Entre des mots ou entre des divisions de niveau supérieur.

> Contrairement à l'apparence, cette balise marque un début et non la fin d'une page. Elle est par conséquent placée au début de la première page et est absente à la fin de la dernière.

Contenu

Elément vide.

Attributs

- **n** : indique le numéro de la page ou du folio.

<q>

Balise du discours direct.

Position

A l'intérieur des paragraphes. Si le discours direct dépasse un paragraphe, le mécanisme de « fragmentation et reconstitution d'éléments virtuels » est utilisé (cf. section 3).

Contenu

Eventuellement, des segments ponctuables **<seg type="sp">** ;
Mots **<w>** et autres éléments de ce niveau.

Attributs

- **`type`** : type de discours direct (**`"spoken"`** par défaut, **`"written"`** ou autre éventuellement).

Exemple

```
A quoy il respondy , sans dire wy , ou nanil , <q> " par foy &slong;ire il
souffira assez se je voy ce qui se fera , " </q> et ce disoit le conte…
```

<seg>

Elément générique pour la segmentation linguistique du texte.

Position

A l'intérieur des paragraphes.

Contenu

Mots, marques de ponctuation, etc.

Attributs

- **`type`** : type de segment (**`"sp"`** pour « segment ponctuable »).
  La limite de **`<seg type="sp">`** et placée avant la marque de ponctuation. La marque de ponctuation est ainsi intégrée à l'unité textuelle qui la suit.

Exemple

```
<seg type="sp" ana="#d9 #z4"> … vouldroy je parler </seg> <seg type="sp"
ana="#c4 #c3">, C'est de chace … </seg>
```

<space>

Espace blanc de taille significative.

Position dans la structure

Entre des mots.

Contenu

Elément vide.

Attributs

- **`dim`** : dimension (**`"vertical"`** ou **`"horizontal"`**) ;
- **`unit`** : unité de mesure (**`"chars"`**, **`"lines"`**, etc.) ;
- **`quantity`** : valeur numérique ;
- **`resp`** : responsable du balisage.

<subst>

Regroupe les éléments **`<del>`** et **`<add>`** en cas de correction scribale.

Position

Représentation imitative.

Contenu

<del> et **<add>**.

<supplied>

Lettres, mots ou groupes de mots ajoutés par l'éditeur ou par un relecteur/encodeur pour corriger les coquilles du manuscrit.

Position et contenu

Représentation normalisée, cf. section 3 ci-dessus.

<unclear>

Lettres, mots ou groupes de mots difficilement lisibles dans le manuscrit. Elément défini par la TEI.

Position et contenu

Cf. section 3 ci-dessus.

<w>

Mot lexical ou grammatical. Elément défini par la TEI ; modèle du contenu modifié par le projet Menota ; attributs ajoutés par le projet BFM-MSS.

Position dans la structure

Unité principale de structuration linguistique. Utilisé à l'intérieur de paragraphes **<p>**.

Contenu

Un élément **<choice>** ou l'une des représentations multiples : **<me:norm>**, **<me:dipl>**, **<me:facs>**.

Attributs

- **bfm:aggl** – indique l'absence de blanc avant du mot suivant (valeurs : *elision, simple*) ;
- **bfm:agglCert** – certitude de l'agglutination ;
- **bfm:agglResp** – responsable de l'identification de l'agglutination ;
- **lemma** – référence à un lemme ;
- **ana** – étiquette morpho-syntaxique ;
- **type**.

5. Saisie et correction des transcriptions (syntaxe compacte)

Les transcriptions sont saisies au niveau imitatif, avec des méta-caractères non-XML permettant de générer les niveaux diplomatique et normalisé.
Si la transcription s'appuie sur le texte d'une édition de référence, celui-ci est traité par un script permettant de le « diplomatiser » tout en gardant les informations nécessaires pour le niveau de représentation normalisée.
Même si elles contiennent des annotations non-XML, les transcriptions compactes sont des documents XML validables contre un schéma conforme à la TEI.

Caractères de raccourci, balises simplifiées

Segmentation

+ – agglutination simple ;
+? – agglutination simple « pas sure » ou « petit blanc » ;
' – élision ;
_ – déglutination ;
_? – déglutination « pas sure » ou « petit blanc ».

Abréviations

La portée de l'abréviation (voir la définition plus haut) est marquée par des parenthèses doubles.

Si la résolution de l'abréviation est régulière (automatique à partir de l'association de la marque et des lettres dans la portée de l'abréviation, seule la marque de l'abréviation est donnée (sous la forme d'une entité MUFI).

Syntaxe compacte	Représentation diffractée

`((&et;))`	`<norm></me:norm>`
	`<me:dipl><ex>et</ex></me:dipl>`
	`<me:facs>`
	`  <bfm:mdvAbbr><am>&et;</am></bfm:mdvAbbr>`
	`</me:facs>`

`((o&bar;))`	`<norm>on</me:norm>`
	`<me:dipl>o<ex>n</ex></me:dipl>`
	`<me:facs>`
	`  <bfm:mdvAbbr>o<am>&bar;</am></bfm:mdvAbbr>`
	`</me:facs>`

`…((&m&dblbar;t))`	`<norm>…ment</me:norm>`
	`<me:dipl>…m<ex>en</ex>t</me:dipl>`
	`<me:facs>`
	`  …<bfm:mdvAbbr>m&dblbar;t</bfm:mdvAbbr>`
	`</me:facs>`

Si la résolution n'est pas « automatique », celle-ci est ajoutée, précédée d'une marque de soulignement. Les lettres « restituées » sont placées entre crochets.

Syntaxe compacte	Représentation diffractée

`((e&bar;_e[st]))`	`<me:norm>est</me:norm>`
	`<me:dipl>e<ex>st</ex></me:dipl>`
	`<me:facs>`
	`  <bfm:mdvAbbr>`
	`  e<am>&bar;</am>`
	`  </bfm:mdvAbbr>`
	`</me:facs>`

`((ch&apomod;&rrot;_ch[evalie]r))`	`<me:norm>chevalier</me:norm>`
	`<me:dipl>ch<ex>evalie</ex>r</me:dipl>`
	`<me:facs>`
	`  <bfm:mdvAbbr>`
	`  ch<am>&apomod;</am>&rrot;`
	`  </bfm:mdvAbbr>`
	`</me:facs>`

`((nr&apomod;e_n[ost]re))`	`<me:norm>nostre</me:norm>`
	`<me:dipl>n<ex>ost</ex>re</me:dipl>`
	`<me:facs>`
	`  <bfm:mdvAbbr>`
	`  nr<am>&apomod;</am>e`
	`  </bfm:mdvAbbr>`
	`</me:facs>`

Grandes initiales (lettrines)

Les grandes initiales sont entourées de doubles accolades. La lettre est séparée de ses attributs (nombre de lignes en retrait, taille réelle en nombre de lignes, couleur principale, éventuellement la décoration) par des deux-points :

Syntaxe compacte	Représentation diffractée

`{{Q:2:7:blue}}uant`	`<me:norm>Quant</me:norm>`
	`<me:dipl><hi rend="initiale" >Q</hi>uant</me:dipl>`
	`<me:facs>`
	`  <bfm:lettrine size="2" sizeAct="7" color="blue">`
	`  Q</bfm:lettrine>uant`
	`</me:facs>`

Variantes de lettres

Entités MUFI
(voir la liste complète dans le *Tableau d'encodage des caractères*, section 6)

Lettres à normaliser

Les lettres qui doivent être modifiées dans la représentation normalisée sont précédées de symboles ***** et **#** :

***u** – *u* diplomatique, à transformer en *v*
***v** – *v* diplomatique, à transformer en *u*
***i** – *i* diplomatique, à transformer en *j*
***j** – *j* diplomatique, à transformer en *i*

Le dièse (#) est utilisé devant les minuscules à transformer en majuscules (noms propres, débuts de phrases dans la représentation normalisée.

Si les deux marques se combinent, le dièse est placé avant l'astérisque.

Les diacritiques de la représentation normalisée (*é, ï, ü,* etc.) sont utilisés dans la saisie initiale.

Ponctuation

Les marques de ponctuation sont balisées **<c type="punct">** (le préfixe **bfm:** est ajouté ultérieurement). Les trois représentations sont fournies, séparées par **%** :

```
<punct>.%,%&punctelev;</punct>
```

Selon les cas, une ou deux représentations peuvent être vides :

```
<punct>,%%</punct>
<punct>%,%.</punct>
```

Le premier exemple représente une de ponctuation ajoutée par l'éditeur, qui ne correspond à aucune marque de la source. Le deuxième exemple représente une marque présente dans la source que l'éditeur a choisi d'éliminer dans la représentation normalisée.

Corrections scribales

La syntaxe compacte est applicable pour le balisage des corrections scribales (suppressions et ajouts de mots ou de caractères) qui concernent un mot ou une partie d'un mot. Les balises XML TEI correspondantes sont doivent être utilisées pour des corrections de taille supérieure.
Les marques utilisées pour noter les corrections scribales en syntaxe compacte sont basées sur les recommandations de F. Masai (1950) avec quelques modifications visant à les rendre plus systématiques.
De façon générale, les doubles crochets servent à délimiter le segment concerné par la correction. Une barre oblique indique l'ajout sur la ligne (à la place du texte supprimé ou d'un espace blanc). Une barre oblique inversée ‹ \ › indique l'ajout au-dessus de la ligne. Un tiret signifie une suppression par rature ; un point souscrit (entité de type **&xdotbl;**) signifie une suppression par exponctuation.

Syntaxe compacte	Représentation diffractée

• addition interlinéaire
`[[\aprés/]]`
	`<me:norm>aprés</me:norm>`
	`<me:dipl>apres</me:dipl>`
	`<me:facs>`
	`  <add place="interlinear">`
	`  apres`
	`  </add>`
	`</me:facs>`

• addition sur la ligne
`[[/et\]]`
	`<me:norm>et</me:norm>`
	`<me:dipl>et</me:dipl>`
	`<me:facs>`
	`  <add place="inline">et</add>`
	`</me:facs>`

• addition en marge	`<me:norm>qui</me:norm>`

`[[\\qui//]]`	`<me:dipl>qui</me:dipl>`
	`<me:facs>`
	`  <add place="margin">qui</add>`
	`</me:facs>`

• suppression par rature (lettre encore lisible)
`vi[[- r]]ent`
	`<me:norm>vient</me:norm>`
	`<me:dipl>vient</me:dipl>`
	`<me:facs>`
	`  vi<del rend="line-through">r</del>ent`
	`</me:facs>`

• suppression par rature (lettre illisible)
`vi[[- ]]ent`
	`<me:norm>vient</me:norm>`
	`<me:dipl>vient</me:dipl>`
	`<me:facs>`
	`  vi<del rend="line-through">`
	`  <gap/>`
	`  </del>ent`
	`</me:facs>`

• suppression par exponctuation
`vi[[&rdotbl;]]ent`
	`<me:norm>vient</me:norm>`
	`<me:dipl>vient</me:dipl>`
	`<me:facs>`
	`  vi<del rend="dotbl">&rdotbl;</del>ent`
	`</me:facs>`

• suppression par tout autre procédé (lettre encore lisible)
`vi[[r]]ent`
	`<me:norm>vient</me:norm>`
	`<me:dipl>vient</me:dipl>`
	`<me:facs>`
	`  vi<del>r</del>ent`
	`</me:facs>`

• suppression par tout autre procédé (lettre illisible)
`vi[[ ]]ent`
	`<me:norm>vient</me:norm>`
	`<me:dipl>vient</me:dipl>`
	`<me:facs>`
	`  vi<del><gap/></del>ent`
	`</me:facs>`

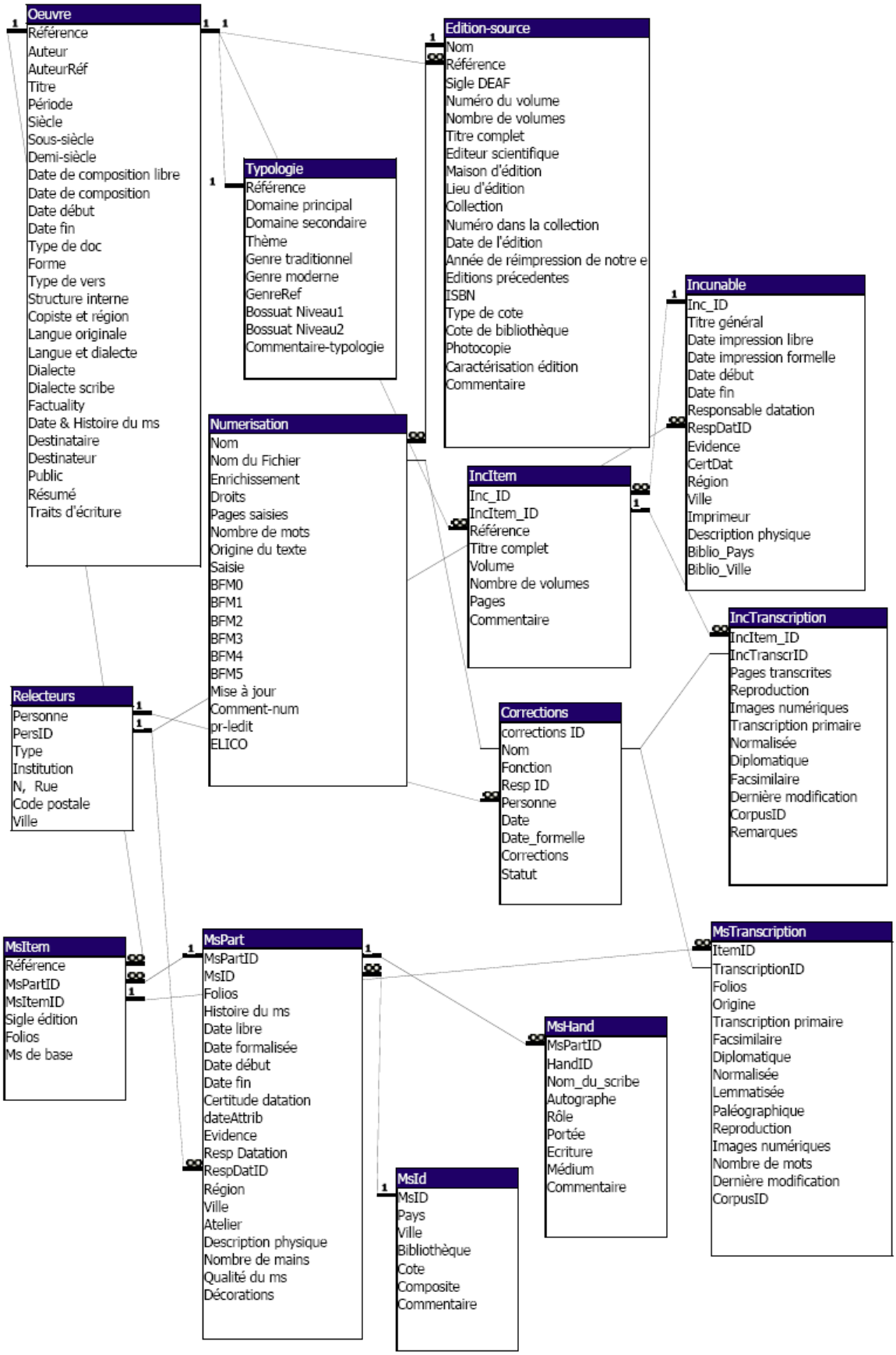
• substitution au dessus d'une rature (lettre encore lisible)
`vi[[- r \ &slong;]]ent`
	`<me:norm>visent</me:norm>`
	`<me:dipl>visent</me:dipl>`
	`<me:facs>`
	`  vi<subst>`
	`  <del rend="line-through">`
	`  r`
	`  </del>`
	`  <add place="interlinear">`
	`  &slong;`
	`  </add>`
	`  </subst>ent`
	`</me:facs>`

• substitution au dessus d'une rature (lettre supprimée illisible)
`vi[[- \ &slong;]]ent`
	`<me:norm>visent</me:norm>`
	`<me:dipl>visent</me:dipl>`
	`<me:facs>`
	`  vi<subst>`
	`  <del rend="line-through">`
	`  <gap/>`
	`  </del>`
	`  <add place="interlinear">`
	`  &slong;`
	`  </add>`
	`  </subst>ent`
	`</me:facs>`

• substitution au dessus d'une exponctuation
`vi[[&rdotbl; \ &slong;]]ent`
	`<me:norm>visent</me:norm>`
	`<me:dipl>visent</me:dipl>`
	`<me:facs>`
	`  vi<subst>`

	<del rend="dotbl">&rdotbl;</del> <add place="interlinear"> &slong; </add> </subst>ent </me:facs>
• substitution par transformation [[e > i]]n	<me:norm>in</me:norm> <me:dipl>in</me:dipl> <me:facs> <subst> <del rend="transform">e</del> <add place="overwrite">i</add> </subst>n </me:facs>
• substitution par superposition d'une lettre écrite sur une autre [[e + i]]n	<me:norm>in</me:norm> <me:dipl>in</me:dipl> <me:facs> <subst> <del rend="unmarked">e</del> <add place="overwrite">i</add> </subst>n </me:facs>
• substitution par superposition d'une lettre écrite au dessus d'une autre (non expressément supprimée) [[e + \ i]]n	<me:norm>in</me:norm> <me:dipl>in</me:dipl> <me:facs> <subst> <del rend="unmarked">e</del> <add place="interlinear">i</add> </subst>n </me:facs>
• substitution sur grattage ou autre type de suppression (lettre supprimée encore lisible) vi[[r / &slong;]]ent	<me:norm>visent</me:norm> <me:dipl>visent</me:dipl> <me:facs> vi<subst> ~~r~~ <add place="overwrite">&slong;</add> </subst>ent </me:facs>
• substitution sur grattage ou autre type de suppression (lettre supprimée illisible) vi[[/ &slong;]]ent	<me:norm>visent</me:norm> <me:dipl>visent</me:dipl> <me:facs> vi<subst> ~~<gap/>~~ <add place="overwrite">&slong;</add> </subst>ent </me:facs>

6. Relations dans la base des descripteurs BFM



7. Tableaux d'encodage des caractères « spéciaux »

Lettres initiales (lettrines)

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
				Encodé au moyen d'éléments

Variantes de caractères

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
ı	= i	ı	0131	*i* simple utilisé dans la transcription, le caractère sans point apparait seulement lors de la visualisation facsimilaire
í	= i	í	00ED	
ð	= d	&drot;	F109	
ɱ	[final m]	&mrdes;	F223 (v2)	
ɲ	&n;	&nrdes;	F228 (v2)	
ɾ	&r;	&rrot;	F20E	
ſ	&s;	&slong;	017F	
ς	&s2;	--		~ &stigma; 03C2
	&s3;	--		→ &slong;
	&s4;	--		→ &slong;
ɵ	s (final)	&sclose;	F128	non utilisé
ẏ	&y-dot;	&ydot;	1E8F	

Ligatures

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
ꝑꝑ	--	&pplig;	EED6	
ꝯt		&ctlig;	EEC5	
ft		&ftlig;	EECB	
ſt		&slongtlig;	FB05	

Caractères exponctués

point souscrit

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
ḍ		&ddotbl;	1EOD	

Abréviations

caractères spéciaux

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
⁊	&et; &et1;	&et;	204A	
⁊̸		&etslash;	F158	
7	&et2;	&ET ;	F142	
	&et3;			
ȝ	--	&etfin;	F155	
÷	&est;	&est;	223B	
ꝰ	&nine;	&usmod;	F151	cf. &us; si sur une lettre
ꝯ	&com;	&condes;	F156	
ſ	--	&is;	F15A	
ꝛ	--	&rum;	F154	

caractères suscrits (combinés)

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
x^{a}	&x[4]-alpha;	x&asup;	0363	
x^{e}	e	x&esup;	0364	
x^{o}	o	x&osup;	0366	
x^{u}	u	x&usup;	0367	
x^{s}	["]	x&ssup;	F027	??? « stigma » suscrit

caractères diacritiques

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
x'	&x-vbar;	x&verbarup;	02C8	
x̄	&x-hbar;	x&bar;	0305	
x̿		*&dblbar;	035E	Absent MUFI, police Cardo
x̃	&x-tilde;	&combtilde;	0303	
x̑		*&combinvbreve;	0311	Absent MUFI, police Cardo
x͡x		*&combdblinvbreve;	0361	Absent MUFI, police Cardo
x̊	&x-omeg;	&ra;	F157	
x’x	&apost;	&apomod;	02BC	
x’x		&combcomma;	0315	code plus correct pour l’apostrophe abbréviatif
x̡	&ier;	&combtildevert;	033E	
x̡	--	&er;	035B	à distinguer de tildevert ?
x̉	--	&ercurl;	F1C8	pb espace après Andron
x̂	--	&ur;	F1C3	^ dans trscr 2001
x̉	--	&urrot;	F153	
x̉	--/ &nine;	&us;	F15B	cf. &usmod;

[4] « x » signifie ici n’importe quel caractère alphabétique.

lettres modifiées

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
ꝺ	d°	&de;	F159	
ꝑ	∥	&pbardes;	E670	
ꝓ	&pro; &pro1;	&pflour;	E67D	
	&pro2;			???
ꝙ		&qslstrok;	E8B1	
ſ	--	&slongslstrok;	E8B8	
ß	--	ß	00DF	
ꝟ	&Ver;	&Vslstrok;	2123	
ꝿ	&ver;	&vdiagstrok;	E8BC (v2)	
	&Vos;			
	&vos;			

Marques de ponctuation et de mise en page

Glyphe	CHR	MUFI	CODE	BFM + Commentaire
·	.	·	00B7	ASCII 183
⁝	&punct;	&punctelev;	F161	
⁞		&punctelevdiag;	F1F0	Est-ce le même que ! ?
⁖		&punctinter;	F160	
⁄	,	/	002F	
ˊ		&virgmin;	F1F7 (v2)	Cf. 05F3
∵		&tridotsdownw;	F1EE	
∴		&lozengedot;	2058	
↯		*&dbloblhyphen;	2E17	Absent MUFI Code 03DF et police Cardo utilisés temporairement
¶		¶	00B6	pied de mouche… faute de mieux
⸌		&parag;	F1E1	crochet adlinéaire
∧		&logand;	2227	marque d'insertion

8. Projets cités :

Charrette : http://www.princeton.edu/~lancelot
Christine de Pizan : http://www.pizan.lib.ed.ac.uk/
Khartês (Université de Liège) : non accessible en ligne
Menota : http://www.menota.org
MUFI : http://gandalf.aksis.uib.no/mufi/

9. Index des éléments (entête)

10. Index des éléments (hors entête)