

# CONSIGNES POUR LE BALISAGE DES TEXTES DE LA BASE DE FRANÇAIS MEDIEVAL

Serge Heiden (slh@ens-lsh.fr)  
Céline Guillot (cguillot@ens-lsh.fr)  
Alexei Lavrentiev (Alexei.Lavrentev@ens-lsh.fr)

CNRS / ENS-LSH, UMR 5191 ICAR



V4.0 2007-12

V3.2 2005-01

V3.1 2004-01

V3.0 2003-02

V2.0 2002-10-30 (intégration du retour des 10 premières relectures)

V1 2002-07

Ce document de travail est élaboré dans le cadre des opérations de relecture et de balisage en XML-TEI des textes de la BFM. Il sert de référence à l'équipe des relecteurs/encodeurs de la BFM et pour des échanges de textes avec nos partenaires.

Ce document est publié librement sur le web à destination de la communauté scientifique dans le cadre de la licence Creative Commons « Paternité-Pas d'Utilisation Commerciale-Partage des Conditions Initiales à l'Identique 2.0 France ». En accord avec cette licence, si vous utilisez ce document dans vos travaux, vous êtes prié de mentionner sa référence (projet BFM, titre, auteurs).



## Sommaire

1	Introduction .....	2
2	Chargement et sauvegarde du fichier .....	2
3	Principes de base .....	3
3.1	Fidélité à l'édition de référence .....	3
3.2	Principes de balisage (cf. Section 1.1.2 du Manuel d'encodage).....	3
4	Délimitation du corps du texte (cf. Section 2.1 du Manuel d'encodage).....	4
5	Structure du corps du texte (cf. Section 2.2 du Manuel d'encodage) .....	5
5.1	Divisions du texte (cf. Section 2.2.1 du Manuel d'encodage) .....	5
5.2	Changement de manuscrit .....	6
5.3	Numérotation des pages (cf. Section 2.2.2 du Manuel d'encodage).....	6
5.3.1	Textes en prose.....	<b>Erreur ! Signet non défini.</b>
5.3.2	Textes en vers.....	<b>Erreur ! Signet non défini.</b>
5.4	Délimitation des unités inférieures (cf. Section 2.2.3 du Manuel d'encodage) .....	7
5.4.1	Textes en prose.....	7

5.4.2	Textes en vers .....	8
6	Mises en évidence typographiques (cf. Section 2.3 du Manuel d'encodage).....	9
6.1	Correction du texte (cf. Section 2.3.2 du Manuel d'encodage) .....	9
6.1.1	Corrections de l'éditeur scientifique (cf. Section 2.3.2.1 du Manuel d'encodage) <b>Erreur ! Signet non défini.</b>	
6.1.2	Corrections des relecteurs (cf. Section 2.3.2.2 du Manuel d'encodage). <b>Erreur ! Signet non défini.</b>	
6.2	Parties manquantes (cf. Section 2.3.2.3 du Manuel d'encodage).....	10
6.3	Délimitation des passages en langue étrangère (cf. section 2.3.3.1 du Manuel d'encodage) .....	10
6.4	Autres mises en évidence (cf. Section 2.3.3.2 du Manuel d'encodage) .....	11
7	Notes ou commentaires (cf. Section 2.3.4 du Manuel d'encodage) .....	12

## 1 Introduction

Le relecteur/encodeur est chargé à la fois de vérifier que le texte qui lui est confié correspond très exactement à l'édition de référence et d'y introduire un certain nombre de balises XML. Il peut éventuellement proposer des corrections aux erreurs manifestes dans l'édition en utilisant des balises appropriées. Elles se présentent sous la forme de méta-informations qu'il faut insérer entre chevrons<sup>1</sup> dans le texte. On s'efforce d'associer une balise unique à chaque type d'information qu'on veut encoder dans le texte.

On utilise pour cela les balises normalisées de la *Text Encoding Initiative* (TEI), consortium qui rassemble à ce jour plus de 80 institutions dans 18 pays<sup>2</sup>. Pour encoder les textes de la BFM nous utilisons un sous-ensemble des balises disponibles dans la TEI. Ce sous-ensemble évolue en fonction des idiosyncrasies rencontrées dans les nouveaux textes traités (il serait bien sûr simpliste de penser pouvoir fixer a priori la structure et le contenu de tous les textes). Par ailleurs, l'usage de certaines balises ou de certains attributs est légèrement adapté du standard TEI pour se conformer aux besoins d'exploitation et d'encodage propres à nos textes. Les balises seront la base des traitements automatiques ultérieurs.

## 2 Chargement et sauvegarde du fichier

Vous recevez, pour une deuxième relecture, un fichier préformaté et pré-balisé (à l'aide des outils de conversion automatique) XML.

Nous recommandons l'usage d'un logiciel de traitement de documents XML spécialisé (par exemple Oxygen) pour toute correction.

En aucun cas le texte ne doit être transformé en un document Word. Aucun enrichissement typographique et aucune mise en évidence Word ne sont admis.

---

<sup>1</sup> < ... >

<sup>2</sup> Le site web de la TEI se trouve à l'URL <http://www.tei-c.org/>.

## 3 Principes de base

### 3.1 Fiabilité des données textuelles

La règle fondamentale de la représentation numérique des textes de la BFM est que la version numérisée doit représenter les données textuelles les plus sûres possibles pour l'étude linguistique de la langue médiévale. Pour des raisons de facilité et de rapidité, les textes de la BFM sont numérisés à partir des meilleures éditions connues de chaque texte médiéval. Dans la mesure du possible, les éditions « bédieristes » (fidèles à un manuscrit de base précis) sont privilégiés. Nous avons pour l'instant pris le parti de **ne pas prendre en compte les notes de l'éditeur**. En revanche, si l'éditeur utilise des marques typographiques à l'intérieur même du texte pour signaler ses interventions (par exemple des crochets pour des mots ou pour des caractères ajoutés), il convient de les baliser de façon à pouvoir accéder, au choix, au texte tel qu'il se présente « avant » et « après » la correction.

Si le relecteur/encodeur désire faire des propositions de correction en cas d'erreur apparente dans l'édition, il peut le faire au moyen de balises spécialisées à cet effet. **En aucun cas il ne doit modifier directement le texte de l'édition de référence ;**

### 3.2 Principes de balisage (cf. Section 1.1.2 du Manuel d'encodage)

- chaque balise est délimitée par des chevrons (< ><sup>3</sup>) ;
- il ne doit pas y avoir d'espace entre le nom de la balise et le premier chevron '<' ;
- les balises permettent de marquer explicitement la structuration du texte (les divisions de toutes sortes : prologue, chapitre, saut de page...) ;
- on distingue les simples éléments frontières (ex : le saut de ligne marque la frontière entre deux lignes), et les éléments emboîtés qui contiennent d'autres éléments (ex : le chapitre contient une portion du texte et il est susceptible d'être divisé en paragraphes...) :
  - on n'utilise que trois balises du premier type<sup>4</sup>. Elles marquent une nouvelle ligne ou un nouveau vers (<lb/><sup>5</sup>), une nouvelle page (<pb/><sup>6</sup>) et un passage manquant dans un manuscrit (<gap/>) ;
  - dans le second cas, le balisage comporte un élément ouvrant et un élément fermant qui comporte un slash en préfixe (ex : <div>...</div>) ;
- certaines balises comprennent, outre l'élément ouvrant et fermant, un attribut particulier auquel on affecte une valeur (ex : pour délimiter la laisse on utilise la balise <ab> comportant un attribut **type** encodant le type de division « laisse » : <ab rend="laisse">...</ab>).
- l'ordre dans lequel on indique les attributs est libre ;
- le nombre d'espaces ou de tabulations situés entre deux mots ou balises dans le corps du texte n'a pas d'interprétation particulière ;
- de même pour les sauts de ligne.

---

<sup>3</sup> Dans tous les exemples qui suivent, les balises sont indiquées dans une police grasse afin de faciliter la lecture. Dans le corps des textes, elles ne doivent bien sûr porter aucun enrichissement Word particulier.

<sup>4</sup> En XML les balises de ce type comportent un slash en suffixe.

<sup>5</sup> pour « line-break ».

<sup>6</sup> pour « page-break ».

## 4 Délimitation du corps du texte (cf. Section 2.1 du Manuel d'encodage)

- Si le texte présente un prologue, un appendice ou tout autre élément distinct du corps même du texte, il est nécessaire de distinguer chacun de ces éléments ;
- l'élément `<text>` délimite l'unité textuelle dans sa globalité (y compris le prologue ou l'appendice) ;
- l'élément `<body>` marque le début et la fin du corps du texte proprement dit ;
- l'élément `<front>` enchâsse les informations liminaires : prologue, sommaire...et surtout le titre de l'œuvre s'il figure sur la première page du corps du texte (on ne tient pas compte des informations et du titre qui se trouvent sur la page de garde) ;
- l'élément `<back>` enchâsse les informations finales : appendice, index... et surtout l'explicit de l'œuvre.

NOTA BENE : Les rares explicits ne se trouvant pas en fin de document sont à placer dans une division `<div>` dont l'attribut `type` a la valeur `"explicit"`.

La structure potentielle d'un texte est donc toujours la suivante :

```
<text>
<front>
<div type="Prologue">
<head>Prologue</head>
<p>
contenu du prologue
</p>
</div>
</front>
<body>
<p>
contenu du texte
</p>
</body>
<back>
<div type="appendice">
<p>
contenu de l'appendice
</p>
</div>
</back>
</text>
```

ou

```
<text>
<front>
<div type="titre">
<p>
titre de l'oeuvre
</p>
</div>
</front>
```

```

<body>
<p>
contenu du texte
</p>
</body>
<back>
<div type="explicit">
<p>
explicit
</p>
</div>
</back>
</text>

```

*Exemple :*

```

<text>
<body>
<ab>
Buona pulcella fut Eulalia,
Bel auret corps, bellezour anima.
Uoldrent la ueintre li Deo inimi,
Uoldrent la faire diaule seruir.
Elle no.nt eskoltet les mals conselliers,
Qu'elle Deo raneiet chi maent sus en ciel.
...
Tuit oram que por nos degnet preier
Qued auisset de nos Christus mercit
Post la mort et a lui nos laist uenir
Par souue clementia.
</ab>
</body>
</text>

```

## 5 Structure du corps du texte (cf. Section 2.2 du Manuel d'encodage)

### 5.1 Divisions du texte (cf. Section 2.2.1 du Manuel d'encodage)

- Le corps du texte (son « body ») peut être structuré en divisions multiples, qui toutes seront indiquées au moyen des balises appropriées ;
- toutes les balises comportent un élément invariant **div** ;
- elles ont toutes également un **type** qui précise la nature de la division (livre, chapitre, chanson...) et un attribut **n** (numéro de chapitre...), ces deux indications étant obligatoirement indiquées (si on les possède).

Exemple :

```
<div type="chapitre" n="1">...</div>
```

```
<div type="livre" n="1">...</div>
```

- si la division comporte un titre, nous vous demandons de l'indiquer au moyen de la balise `<head>`.

Exemple :

```
<div type="chapitre" n="1">
<head>Ci commence li premiers chapitres qui parole de l'office as
baillis.</head>
<p>Tout soit il ainsi qu'il n'ait pas en nous toutes les
graces qui doivent estre en homme qui s'entremet de baillie,
pour ce ne lerons nous pas a traitier premiers en cest
chapitre de l'estat et de l'office as baillis, et dirons briement
une partie des vertus qu'il doivent avoir, et comment il se
doivent maintenir, si que cil qui s'entremetront de l'office
i puissent prendre aucune essample...</p>
</div >
```

### Pseudo-titres

Il peut arriver qu'on rencontre dans un texte un titre qui ne correspond à aucune division logique (et qui n'est pas numéroté). On considèrera alors qu'on a affaire à une « pseudo-division » qu'on balisera au moyen de la balise `<div type="pseudo-div">`. Le titre se trouvera comme dans les cas précédents dans son `<head>`.

## 5.2 Changement de manuscrit

Si l'édition de référence marque le changement de manuscrit, il faut l'indiquer dans une note de début et dans une autre note à la fin du changement (cf. section 8 ci-dessous).

```
<lb/>qui vos devoit faire oblïer ?
<lb/><note resp="proofreader">Passage au ms. B</note>Ge vos ai
fait molt lait servise,
<lb/>car par mon fait estes ocise ;
<lb/>la sorciere dut enchanter
<lb/>par coi deüstes oblïer.
<lb/><note resp="proofreader">Retour au ms. A</note> ...
```

NOTA BENE : Le texte de la note est libre.

## 5.3 Numérotation des pages (cf. Section 2.2.2 du Manuel d'encodage)

- la pagination est déjà indiquée au moyen de la balise `<pb/>` ;
- nous vous demandons de vérifier que la balise se trouve à la bonne place ;
- si ce n'est pas le cas, il convient de placer la balise `<pb/>` entre le dernier mot de la page qui finit et le premier mot de la page qui suit ;
- si les balises sont numérotées, il convient de vérifier l'exactitude du numéro ;
- attention aux illustrations qui peuvent générer des erreurs de pagination, nous vous demandons de nous les signaler (cf. NOTA BENE de la section 2.2.2 du Manuel).

*Exemple :*

```

...
- Or i voist donc, fait ele, car se il demain ne deust
revenir il n'i alast hui par ma volenté.» Et il monte
et la damoisele ausi,
<pb n="2"/>
si se partent de laienz sanz
autre congié, et sanz plus de compaignie, fors
solement dui escuier qui avec la damoisele
estoient venuz. Et quant il sont issu de Kamaalot
...

```

NOTA BENE : Cas d'un mot qui commence sur une page et finit sur une autre :

- Nous vous demandons de supprimer la césure et de mettre le mot tout entier dans la page où il commence.

## **5.4 Délimitation des unités inférieures (cf. Section 2.2.3 du Manuel d'encodage)**

### **5.4.1 Textes en prose**

#### **paragraphe**

- pour les textes dans lesquels le paragraphe est noté explicitement par la balise `<p>...</p>`, nous vous demandons de vérifier que l'élément ouvrant et l'élément fermant sont à la bonne place (il est inutile de reporter le numéro du paragraphe) ;
- soyez particulièrement vigilant à la fin de la page, il peut y avoir une balise de fin de paragraphe qui ne correspond pas à une fin de paragraphe dans l'édition. Il faut l'éliminer ;
- pour les textes dans lesquels les paragraphes sont notés implicitement (par un alinéa ou un saut de ligne) ou ne sont pas indiqués du tout, nous vous demandons de les marquer de manière explicite au moyen de la balise `<p>...</p>` ;

#### **ligne**

- lorsque les sauts de ligne sont indiqués au moyen des balises `<lb/>`, on doit vérifier leur place.

*Exemple :*

```
<p>
<lb/>«Or i voist donc, fait ele, car se il demain ne deust
<lb/>revenir il n'i alast hui par ma volenté.» Et il monte
<pb/>
<lb/>et la damoisele ausi, si se partent de laienz sanz
<lb/>autre congié, et sanz plus de compaignie, fors
<lb/>solement dui escuier qui avec la damoisele
<lb/>estoient venuz. Et quant il sont issu de Kamaalot
</p>
```

## 5.4.2 Textes en vers

### strophe, laisse, refrain...

- pour les textes dans lesquels la strophe, la laisse, le refrain... est marqué explicitement par la balise `<ab type="strophe">...</ab>`, nous vous demandons de vérifier que l'élément ouvrant et l'élément fermant sont à la bonne place (il est inutile de reporter le numéro de la strophe...);
- pour les textes dans lesquels la strophe (ou autre groupe de vers) est notée implicitement (par un alinéa ou un saut de ligne) ou absente totalement du texte, nous vous demandons de l'indiquer de manière explicite au moyen de la balise `<ab type="gv">...</ab>` (il est inutile d'indiquer l'attribut `n` qui comporte le numéro du groupe de vers)

### vers

- nous vous demandons de vérifier que la balise `<lb/>` se trouve bien au début de chaque vers.

*Exemple :*

```
<text>
<body>
<p type="gv">
<lb/>En l'ost n'orent pas lor seignor ;
<lb/>en l'andemain matin al jor
...
<lb/>que il avoit si tost perdu ;
<lb/>molt l'en estoit mal avenu.
</p>
...
</body>
</text>
```

## Numérotation erronée

En cas d'erreur (répétition/suppression) dans la numérotation des vers ou d'une numérotation non standard dans l'édition (des numéros avec des lettres, par exemple), il faut indiquer « à la main » tous les numéros qui ne correspondent pas à la formule : « n+1 » (où n est le numéro du vers précédent).

## 6 Mises en évidence typographiques (cf. Section 2.3 du Manuel d'encodage)

Nous demandons au relecteur/encodeur de rendre compte des différentes marques typographiques insérées dans le texte par l'éditeur scientifique.

La TEI étant avant tout destinée à représenter les notions sémantiques nécessaires à l'analyse du texte, nous demandons au relecteur/encodeur de signaler ces marques typographiques au moyen de balises qui en décrivent la signification.

Comme les marques typographiques varient considérablement selon les éditions dans leur forme et leur signification, il est nécessaire que nous ayons accès à la fois à la signification attachée à chacune de ces marques et à leur forme respective.

Nous demandons donc au relecteur d'analyser l'usage des marques typographiques dans le texte et de remplir le questionnaire *Usage des marques typographiques dans les éditions des textes de la BFM* qu'on lui fournit. Ce questionnaire donne des indications sur les deux cas de figure qui peuvent se présenter :

- l'usage des marques typographiques est homogène (par exemple, les crochets sont toujours utilisés pour marquer les insertions). Il faut alors utiliser une balise qui décrit la signification de chaque marque typographique (les indications fournies par le questionnaire nous permettant de rétablir les marques typographiques utilisées) ;
- l'usage des marques typographiques n'est pas tout à fait homogène. Il convient alors de préciser la forme de la marque typographique utilisée. On adjoint pour cela l'attribut **rend** à la balise qu'on a choisie pour décrire la signification de cette marque.

Les marques de mise en évidence de l'éditeur scientifique sont de deux types :

- marques typographiques au sens propre (italiques, petites majuscules, gras...)
- symboles insérés dans le texte (crochets, parenthèses, points de suspension...)

Le format des textes de la BFM étant le texte brut, il est impératif de baliser conformément à notre usage de la TEI ce que l'éditeur met en évidence dans le texte au moyen des marques typographiques propres (à l'exception des majuscules simples). Les symboles et les majuscules peuvent quant à eux être conservés tels quels dans le texte. Les balises qui encodent leur signification seront par la suite insérées par nos soins grâce aux indications fournies par les réponses au questionnaire.

### 6.1 Correction du texte (cf. Section 2.3.2 du Manuel d'encodage)

Voir la section 2.3.2 du Manuel d'encodage.

Le système TEI dispose d'un certain nombre de balises pour marquer et gérer les interventions éditoriales aux différentes étapes de la rédaction d'un texte. Pour les textes de la BFM il convient de distinguer en premier lieu les corrections que l'éditeur scientifique apporte à son manuscrit de base et qu'il met en évidence par des marques typographiques, et les corrections des erreurs de l'édition proposées par les relecteurs/encodeurs de la version numérisée.

## 6.2 Parties manquantes (cf. Section 2.3.2.3 du Manuel d'encodage)

Les parties manquantes de manuscrits sont à indiquer à l'aide d'un élément `<gap/>`. Cette information est donnée par l'édition de référence. Le responsable de la balise est donc l'éditeur scientifique (`resp="editor"`).

*Exemple :*

```
<lb/>Si angoissés <gap resp="editor" rend="points"/>
<lb/>Que je ne pooie dormir,
```

*Glose :* D'après l'éditeur scientifique, il manque la fin du vers.

## 6.3 Délimitation des passages en langue étrangère (cf. section 2.3.3.1 du Manuel d'encodage)

Si vous rencontrez un passage en langue étrangère (en latin le plus souvent), nous vous demandons de l'enchâsser dans le texte au moyen de la balise `<foreign>...</foreign>`. S'il s'agit d'une autre langue que le latin, il convient de l'indiquer dans l'attribut `lang`. Si l'usage de marques typographiques qui servent à mettre en évidence les passages en langue étrangère n'est pas systématique, il convient de préciser la marque typographique utilisée dans chaque cas à l'aide d'un attribut `rend`. Si le passage n'est pas marqué typographiquement, il faut utiliser `rend="non"`.

*Exemple :*

```
...
- Et tout li haut
homme, et clerc et lai et petit et grant, demenerent
si grant goie a l'esmouvoir que onques encore si faite
goie ne si fais estoires ne fu veus ne oï; et si fisent
li pelerin monter as castiaus des nes tous les
prestres et les clers qui canterent <foreign rend="ital"> Veni
creator spiritus </foreign>
Et trestout et grant et petit plorerent de pec et de
le grant goie qu'i eurent
...
```

## 6.4 Autres mises en évidence (cf. Section 2.3.3.2 du Manuel d'encodage)

Dans le cas où une marque typographique a un autre usage que ceux qui viennent d'être décrits, nous vous demandons de baliser le passage mis en évidence au moyen de l'élément `<hi>`, d'indiquer la forme de la mise en évidence typographique (italiques, majuscules, etc.) dans l'attribut `rend`, et d'ajouter une `<note>` après le passage balisé qui décrit si possible la fonction de cette mise en évidence typographique.

Les valeurs proposées dans la colonne intitulée "fonction" du questionnaire *Usage des marques typographiques dans les éditions des textes de la BFM* peuvent guider la rédaction de cette note.

Exemples :

Texte de l'édition	Texte balisé
si com Tulles le determine ou livre qu'il fist <i>de Viellece</i> , qu'il loe et veust plus que Jeunece,	si com Tulles le determine <code>&lt;lb/&gt;</code> ou livre qu'il fist <code>&lt;hi rend="ital"&gt;</code> de Viellece <code>&lt;/hi&gt;&lt;note&gt;</code> titre d'une autre œuvre <code>&lt;/note&gt;</code> , <code>&lt;lb/&gt;</code> qu'il loe et veust plus que Jeunece,

Texte de l'édition	Texte balisé
Bele, vers cui mes cuers s'acline, RENALS DE BIAUJU molt vos prie Por Diu que ne l'obliés mie.	<code>&lt;lb/&gt;</code> Bele, vers cui mes cuers s'acline, <code>&lt;lb/&gt;&lt;hi hi rend="pmaj"&gt;</code> Renals de Biauju <code>&lt;/hi&gt;&lt;note&gt;</code> nom de l'auteur <code>&lt;/note&gt;</code> molt vos prie <code>&lt;pb/&gt;</code> <code>&lt;lb/&gt;</code> Por Diu que ne l'obliés mie.

Texte de l'édition	Texte balisé
en l'an de Nostre Seignor Dieu Jhesu Crist a mille .iiij <sup>c</sup> x. ans, le .xv <sup>e</sup> . jour dou mois de mars par .j. leundi de la .viiij <sup>e</sup> . indicion.	en l'an de Nostre Seignor Dieu Jhesu Crist a mille .iiij <code>&lt;hi rend="exp"&gt;</code> c <code>&lt;/hi&gt;&lt;note&gt;</code> chiffre en exposant <code>&lt;/note&gt;</code> x. ans, le .xv <code>&lt;hi&gt;</code> e <code>&lt;/hi&gt;&lt;note&gt;</code> chiffre en exposant <code>&lt;/note&gt;</code> . jour dou mois de mars par .j. leundi de la .viiij <code>&lt;hi&gt;</code> e <code>&lt;/hi&gt;&lt;note&gt;</code> chiffre en exposant <code>&lt;/note&gt;</code> . indicion. <code>&lt;/p&gt;</code>

## 7 Notes ou commentaires (cf. Section 2.3.4 du Manuel d'encodage)

L'élément `<note>` est un élément réservé à l'usage du relecteur/encodeur. En effet, il lui permet d'insérer des commentaires dans n'importe quel endroit du texte. Ces commentaires seront alors lisibles par les relecteurs finaux tout en étant clairement marqués comme distincts du texte. En cas de doute lors de l'interprétation d'un passage du texte, nous demandons au relecteur de toujours utiliser un élément `<note>` pour exprimer sa remarque, à l'endroit même où le problème se pose. Il est tout à fait normal de ne pas pouvoir trancher dans certaines situations et toute décision ou non décision doit toujours être explicitement marquée dans les textes. Chaque note doit être "signée" à l'aide de l'attribut `resp`.

*Exemple :*

```
<lb/>A Com cist cheualiers qui ci siet. <note  
resp="proofreader">Deux lettres majuscules initiales</note>  
<lb/>Qu'il ne respont ne un neel.
```

*Glose :* le relecteur/encodeur indique la présence de deux majuscules en début de ligne.