

Structuration et balisage informatiques du glossaire de *Gérard de Nevers*, mise en prose du *Roman de la Violette*

Avant de commencer mon exposé sur le balisage informatique du glossaire de *Gérard de Nevers*, je voudrais vous présenter rapidement l'œuvre en elle-même.

L'Histoire de Gérard de Nevers constitue la mise en prose du *Roman de la Violette* de Gerbert de Montreuil, roman à insertions lyriques rédigé autour des années 1227-1229 qui relate les amours du héros Gérard et de son amie Euryant. Le remaniement en prose a été effectué en milieu bourguignon dans la seconde moitié du XV^e siècle (à une date comprise entre 1451 et 1464). Le but du remaniement en prose est de rajeunir un texte ancien et de l'adapter aux goûts nouveaux.

La mise en prose du *Roman de la Violette* nous est parvenue par deux manuscrits datant de la deuxième moitié du XV^e siècle : le manuscrit B conservé à la Bibliothèque royale de Belgique à Bruxelles sous la cote KBR 9631 et le manuscrit P conservé à la Bibliothèque nationale de France sous la cote BnF fr. 24378.

Je viens de soutenir récemment ma thèse de doctorat qui donne une édition critique du texte en prenant pour base le manuscrit de Bruxelles. **Ma formation de base est donc bien celle d'un littéraire, médiéviste, éditeur de texte, et non pas celle d'un informaticien.**

Dans mon édition, j'ai consacré un volume entier au glossaire, dans lequel je me suis efforcé d'enrichir, de préciser et d'amender le travail effectué par Lowe dans sa précédente édition de *Gérard de Nevers* (1928). J'ai notamment tenté de combler certaines lacunes fâcheuses, comme l'absence de mots caractéristiques du domaine picard, l'absence de locutions typiques du moyen français, ou encore l'absence d'étude détaillée du vocabulaire de la chasse. Par ailleurs, il m'a paru nécessaire d'ajouter les variantes lexicales du manuscrit de Paris dans un complément du glossaire. Cela m'a permis de mettre en lumière les mots ou les locutions absents du manuscrit de Bruxelles, afin de rendre compte de la non fixité du texte, de sa « mouvance », ce qui répond aux souhaits des lexicologues. (J'ai déjà eu l'occasion de parler dans un autre lieu de ce travail et ce n'est pas de cela dont il sera question dans ma communication, mais bien d'aspects plus techniques).

Je m'étais fixé deux objectifs préalablement à la réalisation du glossaire : d'une part composer chaque article sur une **même unité de structure**, et d'autre part rendre les informations contenues dans le glossaire **disponibles pour d'autres travaux**. Je souhaitais en effet pouvoir réutiliser ailleurs les données encodées dans le glossaire, notamment pour l'étude linguistique du texte dans ses particularités géographiques et historiques.

En conséquence, j'ai décidé de constituer mon glossaire de manière informatisée afin de le structurer comme une ressource documentaire, susceptible d'être utilisée comme base de données. Dans cette perspective, j'ai suivi une formation approfondie à l'Université Lille3 dans le domaine du **Traitement Automatique des Corpus**.

1. Le métalangage XML (*Extensible Markup Language* ou « langage extensible de balisage »)

Pour réaliser mon glossaire informatisé, j'ai appris à programmer des outils informatiques et je me suis orienté vers le métalangage **XML** (*Extensible Markup Language* ou « langage extensible de balisage »). Ce métalangage est très utilisé à l'heure actuelle dans le domaine de l'édition et de la création de dictionnaires.

Le principal avantage du XML est de séparer le fond de la forme. Le XML permet en effet de coder des informations sans se soucier dans un premier temps de la présentation future ; la forme est rendue ultérieurement par des manipulations informatiques sur le fichier créé.

Un fichier XML peut être réutilisé selon les besoins : des transformations permettent de produire à partir de lui de nouveaux fichiers réorganisés qui peuvent

-) s'adapter à une grande diversité de **mises en page** (on peut alors jouer à volonté sur le corps, les couleurs, ou la taille des polices)

-) et surtout s'adapter à une grande diversité de **choix éditoriaux**. On peut ainsi sélectionner automatiquement une partie seulement des informations, dans le but de réduire la taille de son document ; ce type de manipulation est particulièrement intéressant dans la perspective d'une édition imprimée qui n'accorde en général qu'une place restreinte au glossaire par rapport au reste du travail d'édition. Il est aussi envisageable de faire disparaître tous les exemples attestés dans le texte pour ne conserver que les descriptions de sens et les références numériques aux sections du texte.

J'ai reproduit ici l'article **boire** tel qu'il figure mis en forme dans le glossaire de *Gérard de Nevers* :

boire {XX, 2}

I. (inf. subst.)

boisson : *Elle, quy savoit a coy le boire touchoit, but tout que riens n'y laissa* {XXVIII, 18}

II. (loc.) **perdre le boire et le mengier**

perdre l'appétit : *Moult pailles et maigres devint, le boire et le mengier perdy ; riens n'estoit quy le peuist assouagier* {XX, 2}

Cet article, choisi parmi les plus simples, comporte un certain nombre de mises en forme visibles (des tailles de caractères différentes, des sauts de ligne, des caractères en romain, en italique ou en gras), mais il contient surtout plusieurs informations balisées grâce au métalangage XML qui enrichissent son contenu, bien qu'elles ne soient pas visibles.

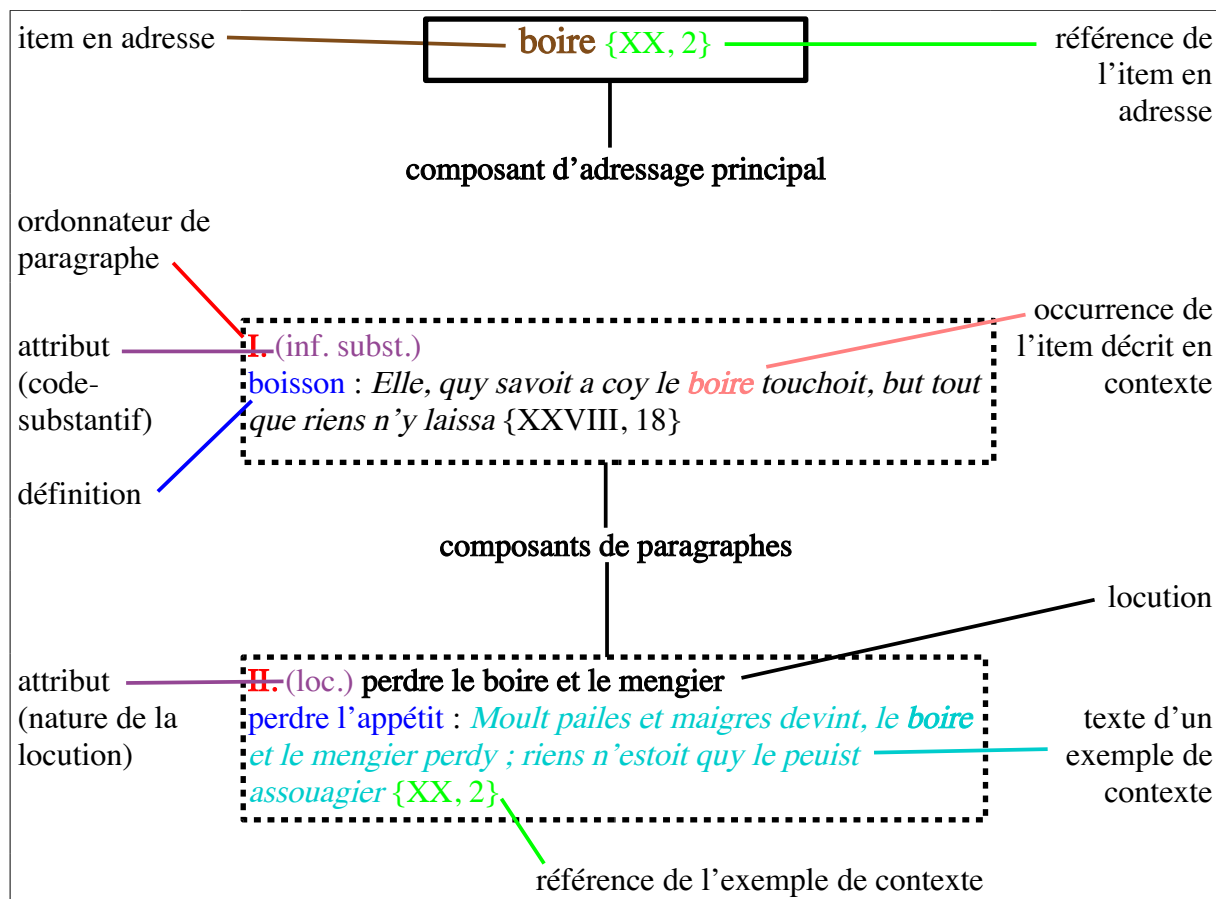


Illustration 1 : Informations balisées dans l'article *boire* du glossaire

Un document structuré en XML ne se réduit donc pas au texte saisi. En effet, le travail de balisage informatique n'est jamais neutre : en répondant aux nécessités de la machine informatique, il oblige à réfléchir sur les besoins du chercheur présent ou futur. Cela exige donc préalablement au codage de l'information une réflexion globale quant aux faits de langue, et une anticipation constante des cas singuliers.

La structuration d'un document avec le métalangage XML permet en effet de créer des contenus qui transforment un corpus en une véritable "base de données" car elle permet d'introduire des informations nouvelles. Ces informations supplémentaires sont de plusieurs sortes : il peut s'agir de la catégorie grammaticale d'un mot ou de la description d'une forme fléchie ; il est également possible de signaler une forme relevée comme étant une variante graphique d'une adresse lemmatisée, ou encore de distinguer les portions de texte qui dans un article représentent des exemples en contexte ou des définitions.

Le texte saisi peut donc être enrichi de plusieurs manières sans que se mêlent les deux types d'informations souhaitées (les données saisies d'une part et le balisage linguistique et informatique d'autre part).

2. À quoi ressemble un document XML ?

La figure suivante reproduit un glossaire minimaliste où figure le seul article **boire** rédigé et balisé en XML :

```
<GLOSSAIRE>
<ARTICLE>
<ARTICLE-STANDARD>
  <ADRESSE-PRINCIPALE>
    <ITEM-ADRESSE>boire</ITEM-ADRESSE>
    <REF-ITEM-ADRESSE>XX, 2</REF-ITEM-ADRESSE>
  </ADRESSE-PRINCIPALE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>I</ORDONNATEUR-DE-PARAGRAPHE>
    <CAT-GRAM>
      <SUBSTANTIF code-substantif="inf. subst."/>
    </CAT-GRAM>
    <BLOC-DEFINITION>
      <DEFINITION>boisson</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Elle, quy savoit a coy le <OCC>boire</OCC> touchoit, but tout que riens n'y
        laissa</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XXVIII, 18</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>II</ORDONNATEUR-DE-PARAGRAPHE>
    <ADRESSE-SUB-LOCUTION>
      <NATURE-LOCUTION/>
      <LOCUTION>perdre le boire et le mengier</LOCUTION>
    </ADRESSE-SUB-LOCUTION>
    <BLOC-DEFINITION>
      <DEFINITION>perdre l'appétit</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Moult pailles et maigres devint, le <OCC>boire</OCC> et le mengier perdy ; riens
        n'estoit quy le peuist assouagier</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XX, 2</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
</ARTICLE-STANDARD>
</ARTICLE>
</GLOSSAIRE>
```

Illustration 2 : Article *boire* rédigé avec le métalangage XML

Pour rappel :

```
boire {XX, 2}
I. (inf. subst.)
boisson : Elle, quy savoit a coy le boire touchoit, but tout que riens n'y laissa {XXVIII, 18}
II. (loc.) perdre le boire et le mengier
perdre l'appétit : Moult pailles et maigres devint, le boire et le mengier perdy ; riens n'estoit
quy le peuist assouagier {XX, 2}
```

Qu'observe-t-on ?

Les informations sont balisées par imbrication des éléments qui contiennent chaque segment de texte.

On repère des portions de texte (comme par exemple « boire » ou « XX, 2 », ou encore « boisson ») et des balises encadrées de chevrons (« < » et « > »), comme par exemple « ARTICLE » ou « BLOC-DEFINITION ».

Un document XML est composé d'éléments. Pour simplifier, chaque élément est constitué : d'une balise ouvrante (par exemple <DEFINITION>), d'un contenu (par exemple « boisson ») et d'une balise fermante (par exemple </DEFINITION>). La barre oblique « / » signifie que l'on a une balise qui ferme l'élément, ce qui est obligatoire en XML.

Dans l'exemple proposé, le document est ouvert par une balise ouvrante « GLOSSAIRE » et se termine par une balise fermante « /GLOSSAIRE ». À l'intérieur de l'élément <GLOSSAIRE>, on trouve un autre élément <ARTICLE> qui contient un élément <ARTICLE-STANDARD> qui lui-même contient un élément <ADRESSE-PRINCIPALE> et deux éléments <PARAGRAPHE>, etc.

```
<-GLOSSAIRE>
<ARTICLE>
<ARTICLE-STANDARD>
  <ADRESSE-PRINCIPALE>
    <ITEM-ADRESSE>boire</ITEM-ADRESSE>
    <REF-ITEM-ADRESSE>XX, 2</REF-ITEM-ADRESSE>
  </ADRESSE-PRINCIPALE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>I</ORDONNATEUR-DE-PARAGRAPHE>
    <CAT-GRAM>
      <SUBSTANTIF code-substantif="inf. subst."/>
    </CAT-GRAM>
    <BLOC-DEFINITION>
      <DEFINITION>boisson</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Elle, qui savoit a coy le <OCC>boire</OCC> touchoit, but tout que riens n'y
        laissa</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XXVIII, 18</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>II</ORDONNATEUR-DE-PARAGRAPHE>
    <ADRESSE-SUB-LOCUTION>
      <NATURE-LOCUTION/>
      <LOCUTION>perdre le boire et le mengier</LOCUTION>
    </ADRESSE-SUB-LOCUTION>
    <BLOC-DEFINITION>
      <DEFINITION>perdre l'appétit</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Moult pailles et maigres devint, le <OCC>boire</OCC> et le mengier perdy ; riens
        n'estoit quy le peuist assouagier</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XX, 2</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
</ARTICLE-STANDARD>
</ARTICLE>
</GLOSSAIRE>
```

Illustration 3 : Article boire rédigé avec le métalangage XML (mise en évidence des éléments)

On relève également dans l'exemple proposé des informations qui sont comprises à l'intérieur de balises : ce sont des **attributs**.

Ils se composent d'un nom et de valeurs selon la syntaxe « nom-attribut="valeur" » (voir par exemple : <SUBSTANTIF code-substantif="inf. subst."/>). Les attributs servent à exprimer des propriétés de l'objet décrit ("fut." pour le futur, "v. intr." pour un verbe intransitif, ou encore "inf. subst." pour un infinitif substantivé). Les valeurs des attributs sont des propriétés fixées dans des listes qui permettent de restreindre le choix de saisie.

J'ai ainsi fixé les valeurs "inf. subst.", "s. fém.", "s. masc." ou "s. masc.-ou-fém." pour l'attribut « @code-substantif ». On peut noter que les attributs enrichissent la description d'un article sans être donnés à voir : dans le texte mis en forme n'apparaissent que les valeurs des attributs (inf.subst.).

```

<GLOSSAIRE>
<ARTICLE>
<ARTICLE-STANDARD>
  <ADRESSE-PRINCIPALE>
    <ITEM-ADRESSE>boire</ITEM-ADRESSE>
    <REF-ITEM-ADRESSE>XX, 2</REF-ITEM-ADRESSE>
  </ADRESSE-PRINCIPALE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>I</ORDONNATEUR-DE-PARAGRAPHE>
    <CAT-GRAM>
      <SUBSTANTIF code-substantif="inf. subst."/>
    </CAT-GRAM>
    <BLOC-DEFINITION>
      <DEFINITION>boisson</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Elle, quy savoit a coy le <OCC>boire</OCC> touchoit, but tout que riens n'y
        laissa</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XXVIII, 18</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
  <PARAGRAPHE>
    <ORDONNATEUR-DE-PARAGRAPHE>II</ORDONNATEUR-DE-PARAGRAPHE>
    <ADRESSE-SUB-LOCUTION>
      <NATURE-LOCUTION/>
      <LOCUTION>perdre le boire et le mengier</LOCUTION>
    </ADRESSE-SUB-LOCUTION>
    <BLOC-DEFINITION>
      <DEFINITION>perdre l'appétit</DEFINITION>
      <EXEMPLE-DE-CONTEXTE>
        <TEXTE>Moult pailles et maigres devint, le <OCC>boire</OCC> et le mengier perdy ; riens
        n'estoit quy le peüst assouagier</TEXTE>
        <REF-EXEMPLE-DE-CONTEXTE>XX, 2</REF-EXEMPLE-DE-CONTEXTE>
      </EXEMPLE-DE-CONTEXTE>
    </BLOC-DEFINITION>
  </PARAGRAPHE>
</ARTICLE-STANDARD>
</ARTICLE>
</GLOSSAIRE>

```

Illustration 4 : Article *boire* rédigé avec le métalangage XML (mise en évidence d'un attribut)

Un document rédigé avec le métalangage XML peut être modélisé **par un arbre**. La figure suivante représente le même glossaire réduit au seul article **boire** dans une structure arborescente :

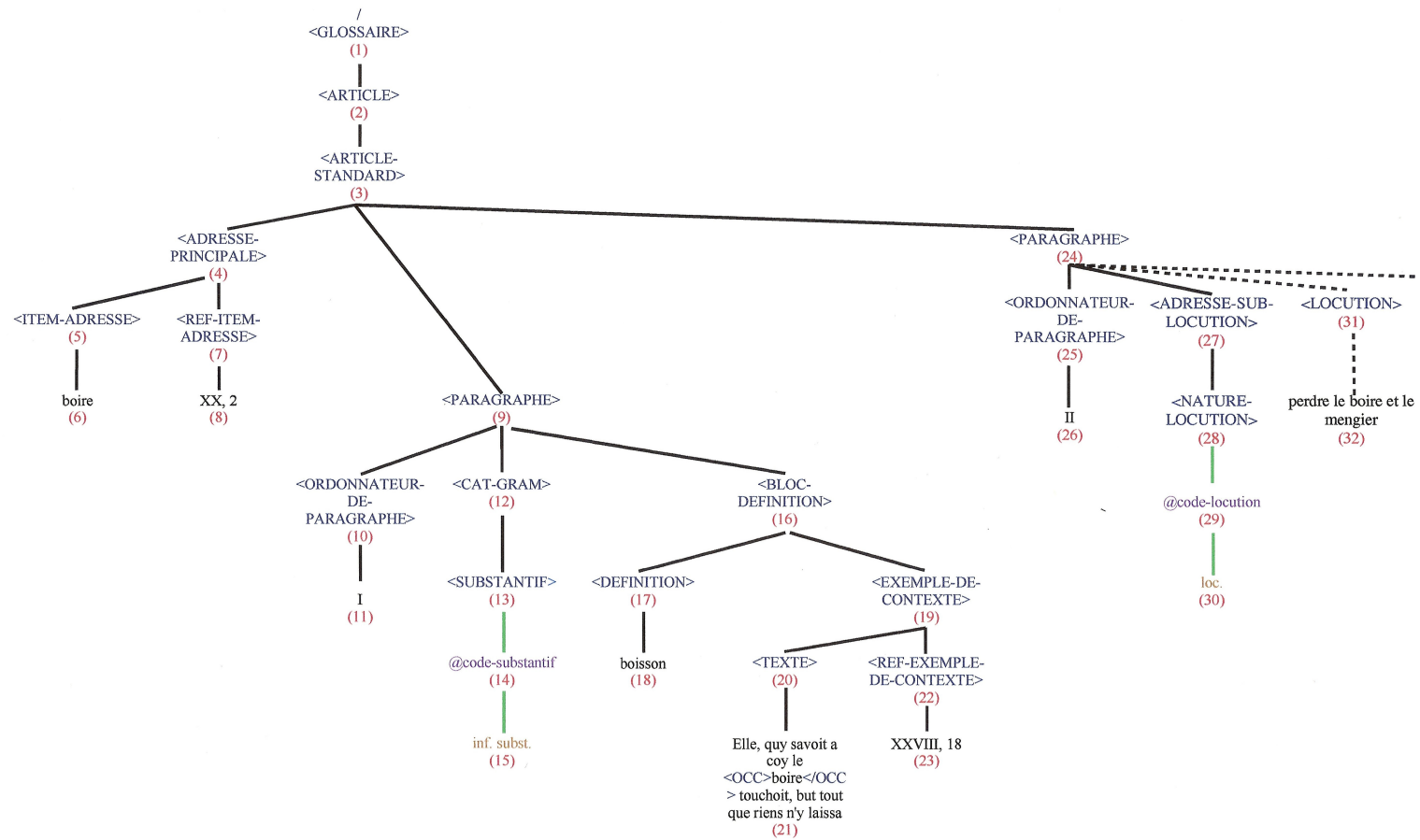


Illustration 5 : Visualisation partielle sous forme d'arbre de l'article *boire*

3. Le rôle de la DTD dans un document XML

Un document XML peut être attaché à une DTD (*Document Type Definition* ou « Définition de type de document ») qui constitue un modèle de structure, un “squelette” unique qui sert à guider la création d’un document. Un document XML qui réfère à une DTD doit toujours lui être conforme pour être validé.

Sans référence à une DTD, un document XML peut être bien formé (c’est-à-dire respecter la syntaxe XML) tout en étant constitué d’un contenu irrégulièrement structuré. Ce cas de figure serait extrêmement préjudiciable dans la perspective d’un glossaire qui recherche précisément **une unité de structure pour l’ensemble des articles qui le composent**. En conséquence, j’ai choisi de lier mon glossaire à une DTD.

On décide dans la DTD de la structure arborescente du document XML. On y fixe le nom des éléments qui servent à coder les informations du document, l’agencement logique de ces derniers les uns par rapport aux autres, leur caractère obligatoire ou facultatif, leur ordre de succession, leur récursivité éventuelle.

La DTD représente ainsi un modèle, un plan où les éléments sont organisés entre eux sous forme d’enchâssements.

La DTD prend la forme d’un arbre qui (pour filer la métaphore) comporte un élément englobant unique à la **racine** (ici, l’élément <GLOSSAIRE>) et des **nœuds** (d’une part les éléments qui correspondent aux subdivisions du document et d’autre part les attributs qui codent certaines données de l’analyse). Les éléments sont imbricables, soit par disjonction, soit par concaténation, ce qui permet d’établir une hiérarchie. Les éléments terminaux de l’arbre (les “**feuilles**”) sont des données de type caractère (du texte) ou les valeurs des attributs (textes pré-renseignés au sein d’une liste fermée).

- À la différence d’un document saisi dans un logiciel de traitement de texte, un fichier XML permet donc d’établir des relations de parenté entre plusieurs éléments. Il y a donc des éléments “parents” et des éléments “enfants” : dans l’exemple proposé, l’élément <ARTICLE> par exemple est le “fils” de l’élément <GLOSSAIRE>. L’élément <ADRESSE-PRINCIPALE> est “frère” de l’élément <PARAGRAPHE>.
- Un élément peut être optionnel (dans la présentation ci-dessous, c’est la signification de l’opérateur « ? » antéposé au nom de certains éléments dans le schéma de la DTD). Dans notre glossaire, les éléments <ORDONNATEUR-DE-PARAGRAPHE> et <ORDONNATEUR-DE-DEFINITION> sont facultatifs. Par exemple, dans le cas d’un article qui ne comprend qu’une seule description de sens (ou « définition ») il n’est pas nécessaire d’introduire un ordonnateur « 1 » qui ne serait suivi d’aucun ordonnateur « 2 ».
- Un élément peut également être répétable (c’est la signification de l’opérateur « + »). Dans notre glossaire, un paragraphe peut contenir plusieurs blocs de définitions ; de même, une définition peut être illustrée par plusieurs exemples de contexte : les éléments <BLOC-DEFINITION> et <EXEMPLE-DE-CONTEXTE> sont donc répétables.
- Enfin, un élément peut être facultatif ou répétable en fonction des besoins (c’est la signification de l’opérateur « * ») : une adresse d’article ne se trouve pas nécessairement sous plusieurs formes graphiques ; par ailleurs, un adjectif masculin

ne se trouve pas nécessairement sous sa forme féminine ; à l'inverse, une même adresse peut comporter de nombreuses variantes graphiques. C'est pourquoi dans notre glossaire, les éléments <VARIANTE-GRAPHIQUE-ITEM-ADRESSE> ou <FORME-FLECHIE> sont facultatifs et répétables.

J'ai reproduit ici une visualisation sous forme d'arbre de la DTD qui a servi de modèle de structure à mon glossaire :

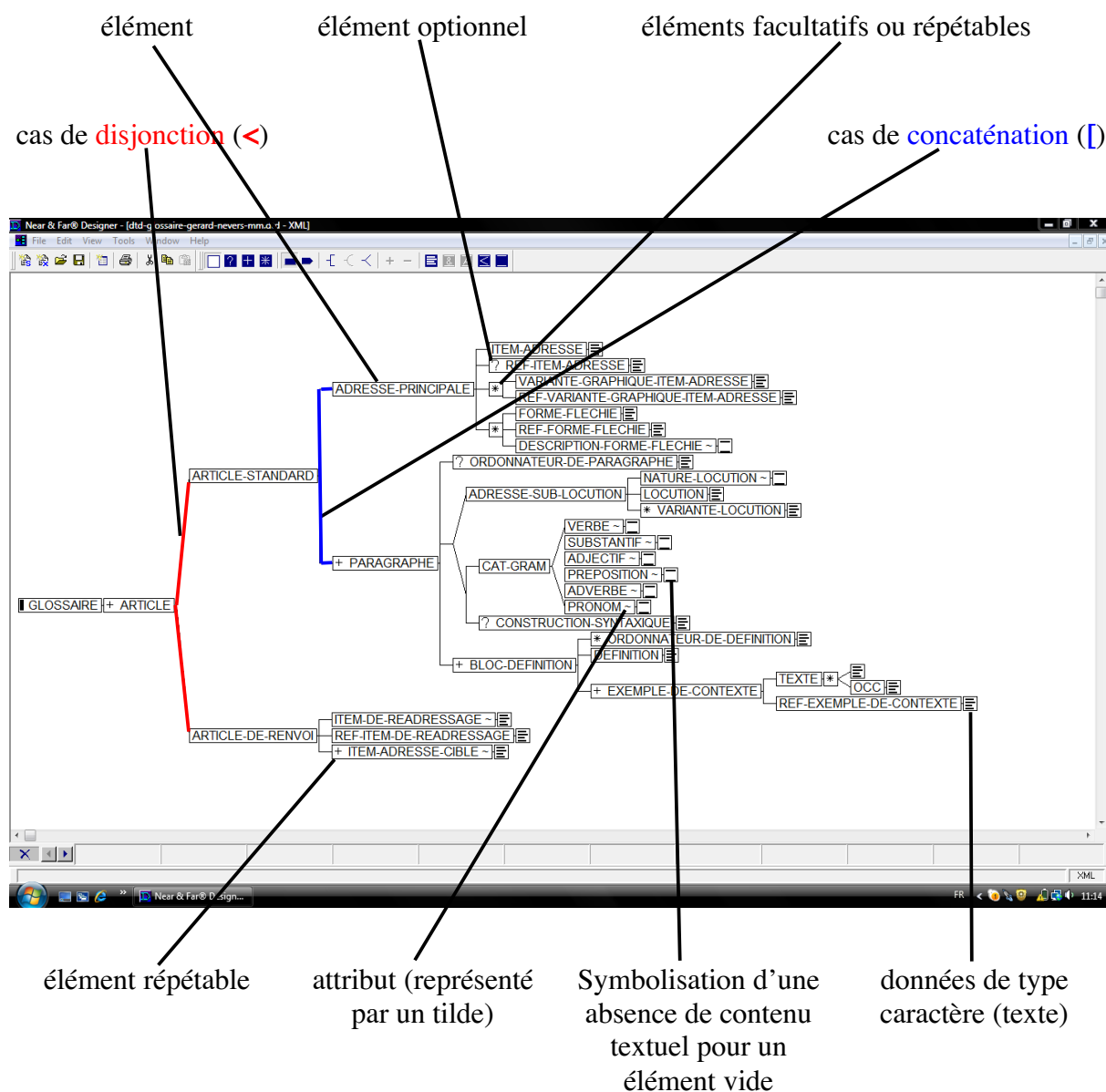


Illustration 6 : Visualisation sous forme d'arbre de la DTD du glossaire de *Gérard de Nevers*

4. Présentation succincte et simplifiée du balisage

Le XML est un métalangage dit « extensible » car il permet de définir de nouveaux éléments, en fonction de la description que l'on veut faire d'un document. **Les noms des éléments ne sont pas fixés par une norme, mais sont créés par le concepteur de la DTD**, ce qui permet une grande souplesse d'utilisation.

Les éléments XML sont donc des **éléments sémantiques**, dont les noms reflètent la **nature** du contenu, par opposition à des éléments prédéfinis qui sont imposés par une norme. J'ai donc fixé les noms des éléments en fonction des informations que je souhaitais voir apparaître dans mon glossaire.

Il serait trop long d'expliquer ici en détails la structuration informatique du glossaire. Je me contenterai de signaler que la programmation informatique a été conçue dans le but d'encoder toutes les informations susceptibles de figurer dans le glossaire d'une édition de texte.

- Le glossaire <GLOSSAIRE> est composé de plusieurs articles (éléments <ARTICLE>).
- Un article se compose soit d'un élément <ARTICLE-STANDARD>, soit d'un élément <ARTICLE-DE-RENVOI> quand la forme observée n'est pas directement identifiable par un lecteur non averti et qu'elle correspond à une forme décrite dans un autre article.
- Toutes les balises qui commencent par *REF* (<REF-EXEMPLE-DE-CONTEXTE>, <REF-FORME-FLECHIE>, <REF-ITEM-ADRESSE>, <REF-ITEM-DE-READRESSAGE>, <REF-VARIANTE-GRAPHIQUE-ITEM-ADRESSE>) donnent pour référence le numéro de section dans laquelle on trouve pour la première fois dans le texte la forme dont il est question.
- Un article standard est constitué d'un élément enchâssant <ARTICLE-STANDARD> qui contient successivement un élément <ADRESSE-PRINCIPALE> dans laquelle seront saisies toutes les informations relatives à l'identification des formes relevées dans le texte, puis un ou plusieurs éléments <PARAGRAPHE> qui contiennent les différents emplois de la forme retenue en adresse.
- L'identification des formes est incluse dans l'élément <ADRESSE-PRINCIPALE>. Dans le cas où l'adresse présente des formes fléchies, on trouve successivement l'élément <FORME-FLECHIE> (qui est facultatif ou répétable selon les besoins) suivi le cas échéant de l'élément <DESCRIPTION-FORME-FLECHIE> qui sert à préciser la nature des formes observées (genre et nombre pour les adjectifs et les substantifs ; mode, temps et personne pour les verbes).
- L'analyse du sens de la forme retenue en adresse est contenue dans l'élément <PARAGRAPHE> qui peut être répétable. La répétition de l'élément <PARAGRAPHE> est motivée par un changement d'emploi de l'unité linguistique sélectionnée (par exemple le passage d'un emploi transitif à un emploi pronominal pour un verbe, ou le passage à la description d'une locution verbale pour un substantif).

La dénomination des éléments et leur agencement s'inspirent en partie du *Dictionnaire du Moyen Français (DMF)* balisé en XML. J'ai repris notamment la disjonction entre articles standards et articles de renvoi et j'ai structuré mes exemples de manière à mettre en exergue les occurrences de l'unité linguistique étudiée en contexte et à les faire suivre de la référence précise dans le texte.

La DTD de mon glossaire ainsi construite a permis d'encoder toutes les informations susceptibles de figurer dans le glossaire d'une édition de texte. Elle a été conçue pour laisser une certaine souplesse rédactionnelle afin que les articles du glossaire soient régulièrement construits et présentent une structure adaptée aux unités linguistiques décrites (substantif, adjectif, verbe).

Toutefois, la DTD de mon glossaire n'a pas la prétention de rivaliser avec celle, plus complexe, du *DMF*. J'ai préféré à plusieurs reprises créer mon propre jeu d'éléments, mieux adaptés à mes besoins, et plus modestes.

5. Les langages associés au XML

Le métalangage XML a favorisé le développement de langages spécifiques destinés à exploiter les contenus des documents XML en vue de leur transformation.

(A) Le langage **Xpath** (*path* pour « chemin ») est un langage qui permet d'exprimer des requêtes sur un fichier XML en localisant les objets recherchés. En d'autres termes, le langage Xpath permet de sélectionner les ancêtres ou les descendants d'un élément.

(B) Le langage Xpath permet de sélectionner des objets (des éléments et leur contenu, des attributs et leur valeur), mais c'est un autre langage qui permet de les transformer : le langage **XSLT** (*eXtended Stylesheet Language – Transformations* ou « Langage XML de feuilles de style en vue d'effectuer des transformations »). Le langage XSLT peut transformer très rapidement un document XML en **produisant un nouveau texte** constitué à partir d'un sous-ensemble des éléments et attributs du document original.

Un certain nombre de requêtes sont envisageables sur l'ensemble du contenu balisé de notre glossaire, ce qui permet de l'utiliser comme une base de données. Grâce aux langages Xpath et XSLT, on peut par exemple compter le nombre d'articles, compter le nombre de verbes ou de substantifs, dresser une liste de tous les articles commençant par la lettre *A*, ou encore isoler toutes les locutions contenues dans les éléments <LOCUTION> d'un article.

Dans la perspective d'une édition de texte, les listes des formes fléchies contenues dans le glossaire permettent par exemple à l'éditeur de **réunir très rapidement des données** sur le texte étudié. L'introduction linguistique de mon travail (l'étude des graphies, des phénomènes phonétiques et morphologiques) a ainsi été grandement facilitée par la réalisation de ces listes qui ont pu mettre en évidence des traits communs entre les formes relevées ou des écarts linguistiques vis-à-vis des normes. Prenons par exemple l'étude morphologique des verbes : j'ai pu étudier plus facilement les radicaux et les désinences verbales des formes fléchies du texte. La programmation en XSLT a donc permis un gain de temps et surtout une efficacité plus grande dans un travail indépendant de l'établissement du glossaire.

Voici deux exemples de listes de formes verbales fléchies obtenues par programmation d'outils d'extraction en langage XSLT. (Les formes verbales fléchies ont été ici classées par ordre alphabétique de l'infinitif et nous les avons fait suivre automatiquement de leur première référence dans le texte et de la forme lemmatisée de l'infinitif).

-) la première permet de mettre en évidence les désinences de première personne du pluriel, quel que soit le mode envisagé (**liste 1**) ;

Liste 1

<i>advisasmes</i> {XLVII, 15} – <i>adviser</i>
<i>irons</i> {XLII, 21} – <i>aler</i>
<i>alasmaes</i> {XXXVIII, 8} – <i>aler</i>
<i>arivasmes</i> {XLVII, 15} - [<i>ariver</i>]
<i>attendons</i> {XVI, 13} - <i>attendre</i>
<i>avons</i> {V, 11} - <i>avoir</i>
<i>arons</i> {LI, 5} - <i>avoir</i>
<i>aviens</i> {VIII, 3} - <i>avoir</i>
<i>ayons</i> {XVI, 6} - <i>avoir</i>
<i>euismes</i> {XXXVIII, 8} - <i>avoir</i>
<i>conseillerons</i> {XIV, 14} – <i>consillier</i>
<i>devons</i> {XLVII, 11} - [<i>devoir</i>]
<i>dirons</i> {XXXIII, 23} - <i>dire</i>
<i>encommenches</i> {XLVII, 14} - <i>encommenchie</i>
<i>enterons</i> {XXIII, 3} – <i>entrer</i>
[...]

-) la seconde permet d'étudier la désinence de première personne du singulier de l'imparfait du subjonctif (**liste 2**).

Liste 2

<i>amaisse</i> {XXVIII, 12} - <i>amer</i>
<i>euisse</i> {III, 5} - <i>avoir</i>
<i>combateisse</i> {XLV, 11} - <i>combatre</i>
<i>cuidasse</i> {XXVIII, 10} - <i>cuidier</i>
<i>deuisse</i> {XVI, 17} - [<i>devoir</i>]
<i>donnaisse</i> {XXVIII, 10} - <i>donner</i>
<i>fuisse</i> {XXXVIII, 20} - <i>estre</i>
<i>feisse</i> {XXXVI, 9} - <i>faire</i>
<i>gardasse</i> {X, 3} - <i>garder</i>
<i>meisse</i> {XXVIII, 11} - <i>mettre</i>
<i>ottroyasse</i> {XXVIII, 11} - <i>ottroyer</i>
<i>osaisse</i> {XXIX, 12} - <i>ozer</i>
<i>peuisse</i> {VII, 2} - <i>poir</i>
<i>promaisse</i> {XLVII, 14} - <i>promettre</i>
<i>saluasse</i> {XXXVI, 9} - [<i>saluer</i>]
<i>seuisse</i> {XLVIII, 9} - [<i>savoir</i>]
<i>veisse</i> {XLVII, 4} - <i>voir</i>
<i>volsisse</i> {III, 1} - <i>voloir</i>

(C) Quand les informations sont encodées dans un document XML, le texte n'est pas mis en forme. Des règles de mise en forme fondées sur la présence des éléments et leur structuration sont établies ultérieurement ; elles servent à produire une grande variété de versions **mises en forme** du texte initialement balisé.

On utilise pour cela le langage **XSL-FO** (*eXtended Stylesheet Language – Formatting Objects* ou « Langage XML de feuilles de style en vue d'effectuer des formatages d'objets »). Le XSL-FO permet de réaliser une mise en page automatique d'un document XML. Il utilise un vocabulaire qui définit des règles de mise en page.

À partir des données initiales contenues dans le document XML, on peut, en appliquant une feuille de style en langage XSL-FO, produire **en série** des pages au format PDF.

La présentation peut donc être pilotée par une feuille de style qui permet d'adapter les informations utiles à une présentation agréable et esthétique tenant compte par exemple des contraintes de format de page, du nombre de colonnes, ou encore de l'espace souhaité entre les colonnes. Il permet par ailleurs de distinguer les mises en page des pages de droite et des pages de gauche, de créer des tableaux, des titres, des lettrines, de produire des pages blanches, des en-têtes ou des pieds de page.

J'ai pour ma part tenté de donner à mon glossaire structuré en XML l'apparence élégante d'un dictionnaire et j'ai pour cela établi une transformation XSL-FO unique traitant tous les articles :

- En pied de page sont introduits les numéros de pages et en en-tête les références des première et dernière adresses d'articles de la page courante.
- Le texte est réparti sur deux colonnes.
- Chaque nouvelle lettre du glossaire est précédée d'une lettrine en corps 100.
- Pour chaque article, l'information est répartie en plusieurs paragraphes. Le premier paragraphe de l'article regroupe l'ensemble des éléments compris dans l'élément <ADRESSE-PRINCIPALE>. Les articles comportent ensuite autant de paragraphes qu'il existe d'éléments <PARAGRAPHE>. À chaque élément <BLOC-DEFINITION> enchâssé dans l'élément <PARAGRAPHE> correspond un sous-paragraphe.
- Les adresses d'article, leurs variantes graphiques éventuelles, ainsi que les formes de renvoi sont mises en évidence en caractère gras et en corps 14 : il s'agit des informations textuelles contenues dans les éléments <ITEM-ADRESSE>, <VARIANTE-GRAPHIQUE-ITEM-ADRESSE>, <ITEM-DE-READRESSAGE> et <ITEM-ADRESSE-CIBLE>.
- Les constructions syntaxiques, les ordonnateurs, les locutions, les variantes de locution et les occurrences dans les exemples sont en gras ; les définitions sont en romain ; les attributs des formes fléchies et les exemples de contexte sont en italique.
- L'affichage des différents attributs a été généré automatiquement, de même que le mot « voir » qui a également été ajouté entre une forme non transparente et sa cible de renvoi, ainsi qu'un certain nombre de marques de ponctuation (des accolades autour des références au texte, des tirets de séparation entre les différentes formes fléchies ou les différents exemples, les deux points après les définitions, un point après les ordonnateurs).

A

aisier {XV, 1} ou **aaysier** {XXXVII, 31}

I. (v. tr.)

1. satisfaire : *vyande avoyent a plenté telle come pour le jour on seüst querre pour corps d'omme aysier* {XXII, 2}
2. pourvoir aux besoins de qq'un, mettre à l'aise : *deux escuiers y laisserent pour Gerart servir et aysier* {XLII, 35}

II. (v. pron.) **soy aysier**

se mettre à l'aise : *Après ce qu'il ot finé sa chanson, vint vers le feu qui estoit moult grant pour soy ressuer et aysier* {XV, 1} - *il s'estoit fait desarmer et aysier au mieulx qu'il avoit peu* {XLVII, 5}

aatine {XLVIII, 4}

(s. fém.)

défi, provocation : *Tant hault monterent les parolles d'un costé et d'autre que par grant aatine ilz ont emprins ung tournoy, ou chascun d'eulx doit estre acompaigniés de ses amys* {XLVIII, 4}

aatir {XXIII, 12} - **aatist** {XLV, 5} *ind. pst. P3*

I. (v. pron.) **soy aatir de**

se vanter de, se faire fort de : *Sy conseille que on voist vers elle et que on luy demande s'elle a commys et fait le murdre ; s'elle dist non, il convendra que Melyatir prengne la lance et l'escu, pour ce que luy meismes s'efforche et aatist de luy prouver le fait contraire* {XLV, 5}

II. (v. pron.) **soy aatir a qq'un**

rivaliser avec qq'un, défier qq'un : *Pensés aultre part, donnés vostre amour ailleurs, car de vous aatir a moy riens ne porriés conquerer !* {XXIII, 12} - *A vous ne me voldroye aatir, mais touteffois je voldroye estre s'ameye* {XXIII, 13}

[abaissier] - abaissoit {LII, 24} *ind. impft. P3* - abaissa {XIV, 15} *ind. p.s. P3*

(v. pron.) **soy abaissier**

1. s'abaisser : *Ainsy come il s'abaissioit pour le lye, prist ung couttel moult trenchant et afilé qu'il avoit a sa chainture* {LII, 24}
2. faire une révérence : *Quant elle vey Gerart, elle prist couleur a muer sy s'abaissa et fu ung pou honteuse, sy dist : « Sire, bon jour vous doinst Dieu. »* {XXXVII, 37}

abandonner {XLVIII, 16} - abandonne {XIX, 16} *ind. pst. P3* - abandonnés {XXX, 8} *ind. pst. P5* - abandonnoit {XV,

3} *ind. impft. P3* - abandonnoyent {XLVI, 6} *ind. impft. P6* - abandonnés {III, 2} *p.p.* - abandonnant {XXIV, 16} *p.pr.*

I. (v. tr.)

1. lâcher : *Gerart [...] haulcha son espee contremont, abandonnant la resne de son destrier* {XXIV, 16}
2. livrer en toute liberté : *Ma terre et mes chasteaulx et tout ce que j'ay au monde vous abandonne pour en faire a vostre plaisir* {XIX, 16} - *Sire, corps et avoir vous abandonne sans nulle villaine pensee* {XXI, 6}

II. (v. pron.) **soy abandonner**

1. se précipiter : *Mais vous gardés sur toute riens que trop ne vous abandonnés : de plus luy serés estrange, de tant serés plus de luy amee* {XXX, 8}
2. se livrer sans retenue, déployer toute son énergie à faire qqch : *Moult fort et asprement se combatirent. Au bien ferir s'abandonnoyent ; il n'y ot celluy qui ne fust navré tant que le sanc leur couroit jusques a l'esperon* {XLVI, 6}

[abandonné] - abandonnee {XXIII, 12} *fém.*

(p.p. emploi adj.)

1. abandonné : *il couvient que seulle et esgaree je le laisse en ceste forest abandonnee aux bestes sauvages* {X, 5}
2. insensé, inconsidéré : *Comment doncques estes vous sy abandonnee ne sy hardye de voloir amer celluy a quy j'ay du tout mon ceur mys !* {XXIII, 12}

abatre {I, 2} - abatoit {XXIV, 20} *ind. impft. P3* - abaty {XXII, 13} *ind. p.s. P3* - abatirent {XVIII, 9} *ind. p.s. P6* - abatu {XXIII, 5} *p.p.* - abatant {XLIX, 17} *p.pr.*

I. (v. tr.)

1. détruire un édifice : *pluiseurs de leurs villes et chasteaux fist abatre et desmolir* {I, 2}
2. faire tomber qq'un violemment, renverser qq'un violemment : *avant ce que sa lance fust rompue ne cassee, il en abaty quatre et du tronchon quy luy demoura ou poing en abaty ung aultre par terre* {XXII, 13} - *Il les ochioit et mehaingnoit : a l'un coppoit bras ou espaulle, il les abatoit et faisoit d'eulx sy grant discipline qu'il n'y avoit celluy sy hardy de l'oser approchier* {XXIV, 20}
3. couper : *Au cheoir qu'il fist, prist sy grant flach que a l'oÿr sambloit ung gros arbre quy fust abatu* {XLII, 15} - *en y mettant toute sa force, assena Liziart sur le healme ung cop sy pesant que l'un des quartiers du healme luy abaty sur l'espaulle, en tel maniere que l'une des oreilles et la joe luy abaty jus* {LII, 18}
4. rabaisser : *moult grant desir ay que son orgoel luy voye abatre* {XVI, 16}

II. (v. pron.) **soy abatre**

se faire tomber : *tant sacherent et boutterent ly ungs l'autre que a terre tous deux s'abatirent* {XVIII, 9}

abatu {XLIX, 16} - abatue {XLIX, 16} *fém.*

(p.p. emploi adj.)

renversé : *Alors le tournoyement commencha : maint chevalier y ot abatu, dont les chevaux couroyent par la prayerye trainant leurs rennes, mainte baniere abatue par terre* {XLIX, 16}

Conclusion

Je viens d'effectuer un tour d'horizon assez rapide et simplifié du balisage de mon glossaire d'édition de texte qui a nécessité un important travail de conception et d'élaboration.

Un glossaire structuré en XML offre à mes yeux de multiples avantages :

-) **c'est une structure ouverte qui permet d'intégrer un nombre non limité d'articles construits selon un même patron.** Mon glossaire comporte ainsi 1500 articles, dont une cinquantaine d'articles de renvoi, il contient environ 2400 formes fléchies, près de 200 variantes graphiques de l'item en adresse (185), environ 900 locutions et plus de 3000 exemples en contexte.

-) **c'est une base de données dynamique susceptible d'être explorée et interrogée**

-) c'est un instrument d'analyse qui rend possible l'**extraction de nombreuses informations sous forme de listes ou de textes, qui constituent des ressources exploitables indépendamment du glossaire** (en particulier pour l'étude linguistique).

-) enfin, c'est un outil éditorial concret et universel : grâce au XML, **le contenu du glossaire peut être réutilisé à souhait**, selon que l'on désire présenter les informations qu'il contient, partiellement ou dans leur intégralité, pour une version imprimée ou électronique, avec la possibilité de suivre plusieurs présentations visuelles.

J'ai ainsi mené une réflexion sur les besoins spécifiques de l'éditeur de texte en moyen français avant de construire de toutes pièces un outil informatique adapté à mes besoins. **Il s'agissait pour moi de faire concorder les exigences des spécialistes en lexicologie, fruits d'une longue tradition, avec les technologies les plus récentes.**

Ce travail ouvre à mon sens de belles perspectives pour les éditeurs de textes, dont les travaux peuvent s'enrichir considérablement grâce aux ressources de l'informatique.