

De l'espace-texte à l'espace virtuel en moyen français

AUTEURS

Darwin SMITH, Charlie MANSFIELD - dans le cadre du projet : ANR EELMAA : (Edition Electronique en Ligne du *Mystère des Actes des Apôtres* de Simon Gréban) Laboratoire de médiévistique occidentale de Paris, CNRS-Université de Paris I Panthéon-Sorbonne (UMR 8589), en ligne à arnoul.vjf.cnrs.fr/actes

SOMMAIRE

Auteurs	1
Sommaire.....	1
Résumé	1
Historique du projet.....	2
De La Grammatologie de nos éléments XML	3
initial.....	4
name	4
Les lien externes.....	5
Autres découpages possibles des données.....	6
Plus de recherches avec le corpus numérique.....	6
Bibliographie.....	8

RESUME

Cet article rend compte des résultats de nos recherches sur le projet consacré aux textes en moyen français dans le cadre du projet EELMAA (Edition Electronique en ligne du *Mystère des Actes des Apôtres* de Simon Gréban) avec le soutien financier de l'ANR (Agence Nationale de la Recherche), dans le cadre d'un programme Blanc (2005).

A notre connaissance, l'informatisation du *Mystère des Actes des Apôtres* est le seul projet d'édition informatique en ligne dans le domaine de la littérature dramatique française pour la période 1350-1500. Ce mystère, composé par Simon Gréban pour le roi René d'Anjou entre 1473 et 1478, est l'œuvre dramatique la plus monumentale que nous possédions dans l'histoire du théâtre occidental (60.000 vers, 494 personnages).

Notre objectif est d'offrir ici un aperçu du chemin que nous avons suivi à l'usage ou à l'information des autres chercheurs qui utilisent l'encodage de textes (le TEI, par exemple, The Text Encoding Initiative, Oxford University), ainsi que les langages de balisage – XML, XSLT, XHTML et CSS – dans le cadre de la linguistique ou des lettres. L'article présente un niveau de développement suffisamment avancé pour que les chercheurs entreprennent leurs propres expériences sans réinventer tout le code JavaScript, et rend également compte de procédés particuliers du DOM (Modèle des objets de Document) créés par Mansfield pour notre projet. Le but ultime du travail sur les textes de ce projet est de rendre des textes et des corpus accessibles en ligne pour les utilisateurs d'Internet et, en même temps, au moyen d'un navigateur (par

exemple Mozilla Firefox version 1.5 et suivante), de leur laisser la possibilité de regarder et étudier la distribution sur la page le texte théâtrale. En somme il s'agit d'une ouverture de l'espace du texte. Le dessein informatique est d'ouvrir un espace numérique pour les chercheurs en littérature, en linguistique, en histoire du théâtre et en paléographie, où ils puissent faire par eux-mêmes des découvertes que le créateur du système informatique pourrait ne pas avoir prévues.

HISTORIQUE DU PROJET

Avant d'aborder la présentation en ligne de ce *Mystère des Actes des Apôtres*, il a été décidé, pour mieux comprendre l'usage des pages web (HTML ou, plutôt DHTML, qui permet d'implanter de l'hypertexte dynamique et JavaScript dans le contenu des pages) pour présenter le processus de performance du théâtre médiéval avec le *Livre de régie* de la Passion de Mons. Cet premier essai avait été mis en ligne sur le serveur du LAMOP pour discuter à l'intérieur de notre équipe de recherche (Groupe d'Etude sur le Théâtre Médiéval, ou GETM). Dès le commencement, notre projet était d'utiliser DHTML afin de démontrer de façon pédagogique les 'fonctionnalités' particulières de la mise en page de cet outil de gestion scénographique.

Veillez voir :

<http://lamop.vjf.cnrs.fr/pages/index.htm>

En mars 2006, Charlie Mansfield est invité à rejoindre l'équipe pour mener la gestion informatique du projet EELMAA. Son introduction d'une gamme restreinte de XML était d'abord conçue pour l'affichage sur l'écran. XML est un langage de balisage extensible qui contient les éléments dans les chevrons < >, éléments qui sont conçus par l'équipe (par chaque équipe) elle-même.

Ce mouvement vers XML donne plus de flexibilité pour le balisage, mais il requiert de former l'équipe à l'usage du langage XML, au langage de transformation XSLT, au langage de DTD (Définition de Type de Document), au langage CSS (Cascading Style Sheets ou feuilles de style en cascade), et enfin de trouver un logiciel pour éditer ces langages. Nous avons trouvé et évalué EditiX, un éditeur XML qui peut éditer ces autres langages d'une entreprise locale (Japisoft 77000 Melun). Nous avons alors commencé le balisage et la transcription, pour lesquels C. Mansfield a écrit les premiers scripts pour le DTD, le JavaScript (ECMAScript, ou ES4), le XSLT et les CSS.

Cependant, en novembre 2007 les universitaires qui gèrent le TEI (Text Encoding Initiative) à Oxford ont sorti une version complète d'XML spécifiquement pour les projets dans les sciences humaines, qui s'appelle : TEI P5.

Un manuel de référence pour chaque élément accessible en ligne à :

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html#>

C. Mansfield, G. Parussa et D. Smith ont rencontré l'équipe de TEI P5 à l'ÉNS, Lyon les 5 et 6 octobre 2006 pendant les deuxièmes journées du Consortium international pour les corpus de français médiéval, et le 15 et 16 novembre 2007 Mansfield organisé les journées de travail pour TEI P5 à Edimbourg. C'est à partir de ce moment que la TEI P5 à commencer à se proposer et à s'imposer comme norme au sein de la communauté universitaire en Europe et les Etats-Unis ; cela signifie pour

nous qu'il faut changer nos balises, et transformer les <éléments> préparés par notre équipe (par chaque équipe) en balises TEI P5 pour que nos fichiers XML (les fichiers XML de chaque équipe) puissent être partagés et nos corpus interopérables.

Veillez voir :

<http://ccfm.ens-lsh.fr>

<http://www.pizan.lib.ed.ac.uk/lab/>

En somme, le lancement de la TEI P5 au cours de la dernière année de notre projet nous a donné l'occasion d'ajouter deux nouvelles dimensions à notre réflexion, soit, (i) le partage du corpus et (ii) l'interrogation du corpus en utilisant les technologies émergentes. De plus, en travaillant depuis mai 2008 avec les langages tels que XSLT et XPath 2.0, sur la suggestion de Naomi Kanaoka, nous avons pu préparer des outils pédagogiques automatisés tels la table d'arborescence pour mieux comprendre et visualiser la structuration d'œuvres dramatiques de grandes dimensions voire monumentales telles que le Mystère des Actes des Apôtres.

Entretemps, est apparue une nouvelle version de JavaScript (ou ES4) :

<http://www.ecmascript.org/es4/spec/evolutionary-programming-tutorial.pdf>

JavaScript ou EC4 est un langage de script (léger), orienté objet et multiplateforme que l'on peut utiliser dans un navigateur comme Mozilla Firefox ; une introduction en français sur les serveurs de Mozilla est disponible à :

<http://developer.mozilla.org/fr/docs/JavaScript>

DE LA GRAMMATOLOGIE DE NOS ELEMENTS XML

Chaque élément XML permet d'associer une donnée, une lettre, un mot, une phrase, voire encore un espace insécable d'un texte transcrit, avec une position (un point) dans l'espace. L'élément balisé est ainsi constitué des parties suivantes :

- le nom de l'élément ;
- pas d'attribut, un attribut ou quelques attributs ;
- les valeurs de ces attributs ;
- le contenu des éléments sauf en cas d'élément prétendu vide.

Voici un élément type :

```
<nom attribut="valeur">contenu</nom>
```

En langage courant on dit que l'élément a 'une balise' d'ouverture et une balise de fermeture.

Dans le cas d'un élément prétendu vide, il faut noter qu'une barre oblique termine alors l'élément sans balise de fermeture :

```
<nom attribut="valeur"/>
```

Maintenant, nous discuterons seulement quelques éléments destinés à donner une vue d'ensemble de la manière dont nous utilisons chacun des langages de programmation pour l'affichage de la transcription. (Sachant qu'il est de fait nécessaire de développer un manuel complet, une bible contenant de tous les éléments utilisés dans les langages de programmation de notre projet, pour ne pas perdre la mémoire permettant la connaissance du travail, même son accessibilité.)

initial

`<initial extent="4">P</initial>`oint n'est besoing icy que je te face,
amy lecteur, quelque longue preface,
mais seulement en brief tu entendras

Cet élément était destiné à définir précisément le nombre de lignes d'une lettrine, tel que nous l'avions conçu antérieurement à notre adoption de TEI P5. Sous P5, l'élément *initial* doit être changé en :

`<hi rend="cap4">P</hi>`

Nous avons alors créé un processus en utilisant XSLT, puis CSS, pour rendre cette balise sur l'écran du navigateur web, d'une façon similaire à ce que serait la lettrine préparée selon les règles typographiques à l'imprimerie.

Dans l'encodage du script de transformation, le fichier XSL ou XSLT, il faut écrire :

```
<xsl:template match="hi [@rend='cap4']">
<span class="cap4">
<xsl:apply-templates/></span>
</xsl:template>
```

Dans l'encodage CSS, il faut écrire :

```
.cap4 {float : left;
margin-top : -9px;
margin-right : 3px;
color : #000000;
font-size : 21mm;
height : 1em;
font-weight : bold;
font-family : Times;
}
```

A l'affichage, cela donne :

P	oint n'est besoing icy que je te face,
	amy lecteur, quelque longue preface,
	mais seulement en brief tu entendras

name

Quand nous rencontrons des noms propres pendant la transcription (il ne s'agit pas ici des noms des personnages, ou noms de rôles), il nous faut ajouter la balise P5 « name » autour du nom propre, avec un attribut où est écrite la forme du nom propre importée de notre « Index des noms propres » - ceci est fait pour empêcher, dans le cas d'une graphie différente d'un même nom, qu'il y ait 'oubli' dans la répertoriatio, ou ambiguïté dans la recherche à un stade ultérieur (dans Wikipedia, ce processus est dénommé « désambiguation »). L'index lui-même est maintenant géré comme un document XML (« fichier noms.xml » sur notre site).

Voici un exemple :

ung frere il eut, `<name ref="arnoul greban">Arnoul Greban</name>`
 nommé,
 gentil ouvrier en pareille science
 et inventeur de grande vehemence ;

En encodage XSLT :

```
<xsl:template match="name">
<a><xsl:attribute
name="href">javascript:popname(' <xsl:value-of
select="@ref"/>')
</xsl:attribute><span class="name" name="noms">
<xsl:apply-templates/></span></a>
</xsl:template>
```

Puis, transformé en HTML :

```
ung frere il eut, <a href="javascript:popname('arnoul
greban')%0A"><span class="name" name="noms">Arnoul
Greban</span></a>
```

Le JavaScript fait fonctionner le lien vers l'*Index des noms propres* :

```
<!-- fenêtre animé en relief MANSFIELD octobre 2007 -->
var newwindow;
function popname(url)
{
var mot="noms.htm#";
urlmot=mot+url;
newwindow=window.open(urlmot, 'name', 'height=400,width=480
, left=380,top=100,toolbar=yes,location=yes,status=yes,men
ubar=yes,scrollbars=yes,resizable=yes');
if (window.focus) {newwindow.focus()}
}
```

LES LIEN EXTERNES

Le laboratoire CNRS d'ATILF nous a permis l'accès au *Dictionnaire du Moyen Français* en ligne via son logiciel de lemmatisation. Cette fois ci encore, il s'agit d'utiliser XML, XSLT, CSS et JavaScript pour réaliser une fenêtre animé (pop-up window) qui contient la lemmatisation et la définition en français moderne de ce que l'utilisateur a marqué avec le curseur :

qui ceux du temps moderne represente,
 car les procès en leurs cueurs si pres ente
 que au monde n'ont volupté si presente,
 de *sufficit* jamais il ne leur chault ;
 larrons des cours d'argent pour appast tente
 qui de piller une lettre patente
 ont obtenu, seellee à mon fer chault ;
 si ton gros cueur, enflé comme ung crapault,
 si ton maling esprit dampnable et cault
 veult declarer de son fait le notice
 secrettement, veulx tu que monte en hault
 à celle fin d'entendre le deffault,
 comme appartient a mon dampnable office ?

LUCIFER
 Tout beau, tout beau, scribe de malefice,
 je ne vueil pas de si pres ton service ;
 auprès de moy ne t'appartient monter.
 Pour monter hault, la divine justice,
 dont il convient que sentence juste ysse,
 en ce bas lac nous a fait **débouter**
 sans à jamais espoir de remonter,
 jugez sans fin souffrir et tourmenter.
 Venez, furies et dyablesses infames,
 que nul de vous ne se vueille absenter,
 penser convient pour le monde tempter,
 plus riens n'avons, tant de hommes que fem

PROSERPINE, MERE DES DYABLES
 Prince mauldicit, tant en fureur t'enflames
 que du chault licit des infermalles flammes



De ce fait, si l'ATILF change ses systèmes, il nous sera possible, à l'avenir, de mettre nos liens à niveau avec une seule reprogrammation (en américain : *leverage*).

AUTRES DECOUPAGES POSSIBLES DES DONNEES

Dès que les textes sont balisés en format XML, les chercheurs peuvent 'découper nos données' selon d'autres formats, en fonction des informations dont ils ont besoin. Le découpage se fait au moyen de deux fichiers XSLT différents qui produisent les pages HTML, une avec une ancre et l'autre avec un nom d'ancre.

Dans le fichier : table.htm

```
<a href="j1critique.htm#JOURNEE I"></a>
```

Pendant que dans le fichier : j1critique.htm

```
<a name="JOURNEE I"></a>
```

```
[IIIa]<a name="Prologue"></a>
```

```
<span class="c">Prologue sur les Actes des Apostres,
```

PLUS DE RECHERCHES AVEC LE CORPUS NUMERIQUE

Maintenant que nous avons le corpus numérique, une prochaine étape pourrait exploiter les balises pour l'exploration du langage. Les recherches de Mansfield (2007, 2008) au niveau doctoral sont dans le champ du langage de mouvement dans l'espace urbain. Au premier niveau nous pouvons faire une table de conjugaison des verbes de mouvement en moyen français, des plus simples – aller, descendre, monter aux plus complexes (débouter), mais pour porter l'enquête plus loin, il faut baliser les répliques de chaque rôle : le chercheur pourrait alors déterminer quel personnage ou

quel type (genre) de personnage utilise les verbes de mouvement : une première ébauche de ce travail à ainsi permis de faire l'hypothèse que les personnages de diables, dont les rôles semblent particulièrement appréciés dans le théâtre de la fin du Moyen Age, ont une utilisation très riche des verbes de mouvement, certainement significative. À première vue, ce balisage semble une tâche ardue, mais comme nous avons déjà balisé les personnages en XML –

pour ensuyvir le sauveur de Nature.

<speaker>Saint Thomas**</speaker>**

Benoistz sont ceulx qui ont creu sans le veoir –

il est facile d'écrire une petite routine (ou autre traitement) pour remplacer chaque balise d'ouverture **<speaker>** avec une balise de réplique (en anglais *speech*) **</sp><sp>**.

Pour l'exemple proposé ci-dessus, cela donne :

pour ensuyvir le sauveur de Nature.

</sp><sp><speaker>Saint Thomas**</speaker>**

Benoistz sont ceulx qui ont creu sans le veoir

[...]

faute de foy faire periller maintz.

</sp><sp><speaker>Saint Bartholemy**</speaker>**

Comme XML fonctionne sur une structure arborescente, une fois ce balisage effectué, les répliques elles-mêmes appartiennent au rôle (ou font corps avec le rôle). De ce fait, on peut ensuite extraire en XSLT ou en XPath 2,0 tout ce que dit le personnage de Lucifer, identifier et analyser ce qui, dans ses répliques, concerne le mouvement dans l'espace : verbes, locutions prépositionnelles, adverbes de position, etc.

Lucifer

Tout beau, tout beau, scribe de malefice,
je ne vueil pas de si **pres** ton service ;
auprés de moy ne t'appartient **monter**.
Pour **monter haut**, la divine justice,
dont il convient que sentence juste ysse,
en ce **bas** lac nous a faict **debouter**
sans à jamais espoir de **remonter**,
jugez sans fin souffrir et tourmenter.
Venez, furies et dyablesses infames,
que nul de vous ne se vueille absenter,
penser convient pour le monde tempter,
plus riens n'avons, tant de hommes que femmes !

C. Mansfield a réalisé un système pédagogique (e-learning) sur ce processus destiné à mettre en évidence les verbes de mouvement pour ses recherches doctorales sur l'espace urbain dans la littérature d'expression française.

BIBLIOGRAPHIE

- Hüe, Denis & Smith, Darwin (2001) *Maistre Pierre Pathelin*, Presses Universitaires de Rennes 2, Rennes. ISBN 2868475477
- Mansfield, Charlie (2008) *Traversing Paris: French Travel Writing Practices in the Late Twentieth Century - An analysis of the work of Annie Ernaux, François Maspéro, Jean Rolin and Anaïk Frantz* Saarbruck, VDM Verlag. ISBN 9783836465038
- Mansfield, Charlie (2007) 'Paris Framed: Twentieth-Century French Writers Crossing the City' pp.175-186 in Bolton, Lucy, Kimber, Gerri, Lewis, Ann and Seabrook, Michael (eds) (2007) *Framed - Essays in French Studies* (Modern French Identities 61) Oxford & Bern, Peter Lang AG ISBN 3039110438
- Mansfield, Charlie (2006) *Europa – English Verse with Parallel French* Paris, Thélès. ISBN 2847767606
- Laidlaw, James & Mansfield, Charlie (2006) 'Designing a Digital Version of British Library Harley MS 4431. The Making of the Queen's Manuscript' in Van Hemelryck, Tania & Van Hoorebeeck, Céline (eds) *L'écrit et le manuscrit à la fin du Moyen Âge* Turnhout, Brepols ISBN 2503519911
- Mansfield, Charlie (2006) 'ABC.HTM - l'écriture numérique' in DalMolin, Eliane & Murphy, Carole (eds) *Contemporary French and Francophone Studies: Sites*, Volume 10 Issue 10.3 : Verbal, Visual, Virtual, London & New York, Routledge. ISSN 1740-9292 University of Connecticut <http://www.sites.uconn.edu/abc.html>
- Smith, Darwin (2002) *Maistre Pierre Pathelin. Le Miroir d'Orgeuil et la traduction de la version inédite du Recueil Bigot, XVe siècle, Mss Paris, B.N.F Fr 1707 & 15080* Tarabuste, Saint-Benoît du Sault. ISBN 2845870299
- Smith, Darwin (1998) 'Les manuscrits « de théâtre » Introduction codicologique à des manuscrits qui n'existent pas', *Gazette du livre médiévale*, no 33, automne 1998, 1-10.