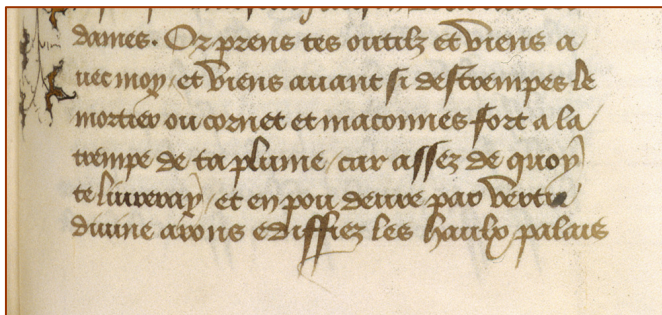


CHARLIE MANSFIELD

2 mars 2007



**Modéliser l'espace dans
les textes
en moyen français :
le développement d'un
Script Informatisé pour
Faire le Tamisage
(SIFT)**

Table de matière

Résumé	1
Le but scientifique : 'tamiser' en modélisant.....	2
Le contexte du travail	2
La fonction du dénouement dans le manuscrit de Mons.....	2
Le développement de la modélisation avec le manuscrit Harley 4431.	4
Une édition diplomatique fait d'une édition savante grâce au DOM	5
Ajouter un caractère à une police	7
Bibliographie	7

Résumé

Cet article rend compte des résultats de nos recherches sur les projets consacrés aux textes en moyen français dans les cadres des deux projets auxquels nous sommes associés pour la période 2005-2008 : les projets ANR (Agence Nationale de la Recherche) à Paris et AHRC (Arts and Humanities Research Council) à Edimbourg. A notre connaissance, ce sont les deux seuls projets d'édition informatique en ligne dans le domaine de la littérature française pour la période 1350-1500. Notre objectif est d'offrir ici un aperçu aux autres chercheurs qui utilisent l'encodage de textes (le TEI, par exemple, The Text Encoding Initiative, Oxford University), et les langages de balisage – XML, XSLT, XHTML – 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 pour ces deux projets. Le but ultime du travail sur les textes de ces deux projets est de rendre 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 changer les couleurs et tailles de police sur leurs écrans (le 'tamisage'). 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 paléographie et en codicologie où ils puissent faire des découvertes eux-mêmes, découvertes que le créateur du système informatique pourrait ne pas avoir prévues, dans un vrai esprit de virtualité (voir : Charlie Mansfield (2000) : « What is the virtual ? »)

Le but scientifique : ‘tamiser’ en modélisant

L’espoir était de pouvoir ‘tamiser’ un texte pour rendre évidents certains éléments du corpus. Pour nos projets, il était très important de fabriquer un outil qui permette aux médiévistes de faire ce tamisage, sans avoir besoin de recourir aux services d’un informaticien à chaque instant. C’est la vitesse du processus de modelage qui donnera aux chercheurs les meilleures chances de pouvoir questionner leurs données : ‘Et si on change cela ... ?’ Donc, Mansfield a créé, en juin 2006, SIFT (tamiser) en utilisant les langages d’Internet, JavaScript et DHTML. SIFT donne à médiéviste un system de modélisation en avant, pendant, et après, ses choix des balises XML pour son corpus.

Le contexte du travail

Un peu d’histoire pour mettre notre travail en contexte. Notre premier défi était centré sur la présentation d’une page d’un livre médiéval de régie théâtrale, dans le cadre de l’*Edition électronique du Mystère des Actes des Apôtres*, de l’Agence Nationale pour la Recherche (ANR, projet NT-NV-31), sous la direction de Darwin Smith au Laboratoire de médiévistique occidentale de Paris, CNRS-Université de Paris I -Panthéon-Sorbonne (UMR 8589).

Le *Mystère des Actes des Apôtres*, 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). Avant d’aborder la présentation en ligne de ce texte, il a été décidé, pour mieux comprendre et faire connaître le processus de performance du théâtre médiéval de présenter le livre de régie de la *Passion* de Mons.

Une grande représentation de la *Passion*, eut lieu à Mons, du 2 au 8 juillet 1501, où furent invités, outre les habitants des villes de la Flandre méridionale, de nombreux personnages de marque, comme le duc de Bourgogne, l’archevêque de Cambrai, les ambassadeurs de nombreux pays. De cette représentation, qui dura 6 jours complets (matin et après-midi, agrémentée de nombreux jeux de rhétorique), il nous reste un livre de compte extraordinairement détaillé et le livre du régisseur de la représentation, appelé « Abrégé », un document unique dans le domaine français, qui nous renseigne précisément sur l’organisation de la mise en scène, aussi bien le jeu des acteurs, que les trucages réalisés pour les « effets spéciaux ». Le texte du mystère était repris de la *Passion* d’Arnoul Gréban, frère de Simon Gréban, et ce texte a servi de base à la quasi-totalité des *Passions* jouées dans le domaine français entre 1455 et 1550.

Par l’expérience de modélisation sur une page du matériel du manuscrit de Mons (voir Figure 2 p.4 SMITH 1998), le défi devient manifeste. Sur la page de Mons, il s’agit de représenter la distribution visuelle du texte autour de la page et en même temps de donner le contrôle de la modélisation au chercheur pour ouvrir le champ d’expérimentation.

La fonction du dénouement dans le manuscrit de Mons

Très vite, il a été compris que la méthode la plus intéressante et la plus claire est d’offrir au chercheur une visualisation de la distribution du texte sur l’espace de la

page du manuscrit. Deuxièmement, la solution offre une méthode pour actualiser l'affichage sur l'écran. De cette manière, le chercheur peut voir instantanément le texte entier du personnage et peut retourner au texte de l'abrégé, avec Ctrl R (actualiser la page).

Voici donc l'encodage de cette fonction, qui utilise JavaScript (ECMAScript). Les fonctions écrites en JavaScript sont notées directement dans la page d'HTML. Elles sont interprétées du côté client (client-side) en temps réel par les navigateurs qui, aujourd'hui possèdent le logiciel d'interprétation correspondant. Ces fonctions ne requièrent ni l'accès aux scripts sur le serveur, ni la modification de la page HTML sur le serveur lui-même. Cependant elles peuvent changer le DOM le « Modèle des objets de Document », et cela modifie la page courante dans l'ordinateur pendant cette session d'observation. Le programmeur ne peut accéder qu'aux éléments de la page HTML qui ont reçu dans leur repère d'ouverture un attribut **name** ou un attribut **id**.

En programmation un variable est un lieu pour stocker une valeur ou un mot un peu comme en algèbre

Veillez voir, par exemple, le repère d'ouverture ci-dessous où le programmeur a créé l'attribut **id** pour le balisage **div** puis lui a donné une valeur de 'ficheDiv2' :

```
<div class="fiche" id="ficheDiv2">
voici ... sa gloire.
</div>
```

Avec la méthode : **document.getElementById** le programmeur peut adresser un objet du document HTML ayant une propriété **id**. Le codage de la fonction JavaScript lui-même réside dans la section **<head>** de la page d'HTML sous la rubrique **<script>**. Au-dessous nous voyons les variables : **ficheText2** et **dit1**, qui sont utilisées pour changer le contenu de la division :

```
function doCheck2()
{ var dit1="Mauvais serpent et decevable, <br />par ta
perverse ambicion <br />recevras malediccion : <br />de
toutes bestes aras pis,";
var ficheText2=dit1+"";
if(document.getElementById)
document.getElementById("ficheDiv2").innerHTML=ficheText2;
}
```

On remarque, dans le code ci-dessus, que dans la phrase stockée au variable : 'dit1' les balises HTML sont ajoutées par le programmeur ; dans cet exemple la balise **
** pour un saut de ligne. Il faut faire tenir compte du fait que le contenu du variable 'dit1' ne subit pas de sauts de ligne dans le code source HTML. Ceci n'est pas apparent dans le document que vous avez entre les mains qui est en format WORD et qui fait aller à la ligne automatiquement.

Une version interactive du livre de régie de la *Passion* de Mons est disponible sur Internet depuis janvier 2007 aux pages de notre laboratoire sous la rubrique « Mons » à :

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

Le développement de la modélisation avec le manuscrit Harley 4431.

Le programmeur prend toutes les balises <name> et les transforme en : pour obtenir un accès à chaque occurrence d'un personnage dans le corpus. Il faut ajouter ici que c'est une tâche minutieuse de baliser un corpus de moyen français.

La Méthode JavaScript : **window.document.getElementsByName** retourne un tableau (Array) d'objets HTML ayant un nom défini dans la propriété *name* de la balise de l'objet. Veuillez remarquer que le mot « Elements » est au pluriel, parce qu'il y a un tableau des éléments, qui commence à [0], comme : doc[0], doc[1], doc [2] et cetera.

En utilisant JavaScript, le programmeur peut changer les attributs de style directement dans le DOM. Il faut se souvenir que le DOM fournit un ensemble standard d'objets pour représenter des documents HTML ; par exemple, on peut faire l'appel des éléments par leur 'Name', en ce cas avec le 'Name' « noms » en utilisant l'appel :

getElementsByName

On écrit alors dans la section <script> de la page HTML :

```
document.getElementsByName("noms")[i].style.color = colour;
```

Mais, il faut créer et puis faire les changements au tableau parce qu'il y a une liste des éléments dans chaque Name.

```
var i=0;  
var huge=9999;  
var array2=document.getElementsByName("noms");  
for (i = 0; i <= huge; i++)  
{ array2[i].style.color = colour;  
array2[i].style.fontSize = mm;
```

C'est dans les deux dernières lignes de code que le programmeur a fait appel (ou changé) deux des propriétés de l'objet qui s'appelle « .style. » : la propriété « .color »

et la propriété « .fontSize. » Un appel à la propriété « .color » change le couleur du texte et un appel à la propriété « .fontSize » change la taille du police du texte dans tous les éléments où la variable Name a le valeur 'noms'(voir exemple ci-dessus).

De cette manière, le programmeur peut donner aux utilisateurs la possibilité de changer les styles au fur et à mesure sans utiliser les styles dans les CSS (Cascading Style Sheets ou feuilles de style). Cet outil permet une étude extrêmement visuelle du manuscrit. Et, comme un géologue, on peut examiner les strates colorées pour détecter les répétitions, et, peut-être, les modèles sous-jacents à la structuration des éléments du corpus.

Pour l'avenir le développement que nous avons planifié vise à ajouter cette deuxième technologie, soit le « getElementByName », à la visualisation de la page du manuscrit de la *Passion* de Mons pour mieux reproduire les effets créés par le copiste médiéval. Puis, avec les leçons tirées de cet essai, aborder la programmation du grand corpus de 60 000 vers du *Mystère des Actes des Apôtres* de Simon Gréban. Mais, dès à présent, ce que nous avons mis au point par l'utilisation de ces nouvelles technologies de DOM et de nouveaux navigateurs, donne l'opportunité d'améliorer nos choix pour le balisage du corpus des œuvres de Christine de Pizan dans le manuscrit Harley 4431 de la British Library et des *Actes des Apôtres* de Simon Gréban.

Une édition diplomatique fait d'une édition savante grâce au DOM

Il est toujours difficile, sinon impossible, de décider à l'avance quel mots baliser au commencement d'un projet de transcription d'un grand corpus textuel (voir, par exemple, Laidlaw & Mansfield 2006) mais avec nos expériences de cette technologie et l'appel au DOM :

`document.getElementById`

nous pouvons désormais trouver des solutions suffisamment puissantes pour tenter de résoudre les difficultés régulièrement rencontrées dans les transcriptions de grands corpus textuels en moyen français. Parmi les difficultés rencontrées, nous citerons la question des caractères accentués en français moderne mais écrits sans accents en moyen français : « é » et « e », et cetera. La savant souvent écrit un accent aigu sur les caractères e de Moyen Français pour rendre le corpus plus lisible par un lecteur moderne, mais en même temps le médiéviste veut garder un copie diplomatique du transcription. Ici la programmation en SIFT peut sembler supprimer les accents de l'édition savante et rendre une édition diplomatique en utilisant le seul texte:

Premièrement le programmeur doit ajouter les balises en début et fin de corpus de Moyen Français en HTML pour donner une identité à la division du texte, en ce cas, le texte entier :

```
<div id="source">
```

...

```
[38a:1] Bon cheualier où tous biens sont compris
[38a:2] noble uaillant et de\ royal ÷lignage
...
```

```
</div>
```

Puis il faut écrire en JavaScript une fonction de 'split', qui trouve le contenu du variable : **hifind[0]** au-dessous. La fonction ici s'appelle **supaccents** parce qu'elle trouve un **é** et puis elle le remplace par **e**.

```
function supaccents()
{
var find="";
hifind[0]=thisform.acc0.value;
var array9="";
var occurs=0;
find = new RegExp(hifind[0], "g");
var replace="";
replace='<span class="accent">' + "e" + '</span>';
array9 = document.getElementById("source");
var col_array=array9.innerHTML.split(hifind[0]);
occurs=col_array.length-1;
count[0].value = occurs;
totaccer = totaccer + occurs;
totacc.value= totaccer;
array9.innerHTML=array9.innerHTML.replace(find, replace);
```

De cette manière, la transcription contient les caractères modernes avec ses accents en édition savante puis sur l'écran l'étudiant peut demander une édition diplomatique, sans accents, en cliquant sur le lien de la fonction JavaScript qui supprime les accents : **supaccents**. Il suffit, donc, de garder un exemplaire unique du corpus, une édition savante avec les accents.

Pour utiliser cette méthode, le 'Mozilla Developer Center' donne les conseils suivants : « innerHTML récupère ou définit tout le balisage et le contenu au sein d'un élément donné [...] Elle ne devrait jamais être utilisée pour écrire des parties de tableaux, les méthodes W3C DOM devant être utilisées à la place, mais elle peut cependant servir à écrire un tableau entier ou le contenu d'une cellule. »

(<http://developer.mozilla.org/fr/docs/DOM:element.innerHTML>)

Nous utilisons ici « find = new RegExp(hifind[0], "g") » pour trouver toutes les caractères (la lettre **é** en l'occurrence), puis « split » pour couper la chaîne de texte et enfin « innerHTML.replace » pour remplacer le **é** avec un simple **e** sans accent.

Un modèle de cette méthode, qui manie un corpus moderne sans balises XML, est disponible sur le site du projet AHRC créé par l'auteur à l'université d'Edimbourg à :

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

avec les remarques et mode d'emploi à :

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

Ajouter un caractère à une police

Si, par exemple, le but est de montrer 4 formes particulières à la lettre M majuscule d'un manuscrit pour une édition savante on peut utiliser mon fonction JavaScript supaccents de au-dessous. Pendant la transcription du manuscrit il faut que le paléographe balise chaque occurrence de la forme du 'M' majuscule par votre propre codage : M1, M2, M3 et M4.

Il faut alors utiliser un logiciel pour dessiner les 4 caractères en questions – appelés M1, M2, M3 et M4 – au moyen, par exemple, de Font-Creator 5 (de la compagnie High-Logic). Ces 4 nouveaux « M » sont définis dans votre propre police en utilisant l'espace à la fin du fichier et les balises XML définies pour chaque « M ».

On peut utiliser le balisage XML TEI P5 pour un glyph, veuillez voir :

<http://www.tei-c.org/release/doc/tei-p5-doc/html/CH.html>

et

<http://www.tei-c.org/release/doc/tei-p5-doc/html/WD.html>

```
<g ref="M1">M</g>
```

La balise **g** est sigle pour *gaiji*, un mot japonais qui signifie un caractère hors-normes.

Puis, en ajoutant l'encodage JavaScript au page du HTML de votre corpus on peut régler l'affichage de M en police standard et puis en police dessiné par un seul click ou Ctrl R.

Auteur : M. Charlie MANSFIELD - dans le cadre des Projets : ANR EELMA : (Edition ELelectronique du Mystère des Actes des apôtres de Simon Gréban) Laboratoire de médiévistique occidentale de Paris, CNRS-Université de Paris I Panthéon-Sorbonne (UMR 8589), et AHRC "The Making of the Queen's Manuscript", University of Edinburgh, Ecosse. Avec remerciements à Darwin SMITH pour les précisions médiévistiques.

Bibliographie

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

Lefebvre, Henri (2000 for 4th edition used here) [original 1974] *La production de l'espace* Paris, Anthropos.

— (1991) Nicholson-Smith, Donald (English translation) *The Production of Space* Oxford, Blackwell.

Mansfield, Charlie (2006) *Europa – English Verse with Parallel French* Paris, Thélès.

— (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

— (2000) What is the virtual - the writings of Pierre Lévy in *The Web Journal of French Media Studies* Volume 3, Number 1, October 2000 [<http://wjfms.ncl.ac.uk/>] ISSN 1460-6550.

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.