

msdescription
transcr
(gaiji)

Guidelines, chap.10-11 (et 5)

Structuration des données et des documents : balisage XML

La TEI et les manuscrits

J.B. Camps – mercredi 29 novembre 2017

*M2 Technologies numériques appliquées à l'histoire
et Humanités numériques*



Ce support de cours est mis à disposition selon
les termes de la licence *Creative Commons*
Paternité – Pas d'utilisation commerciale –
Partage à l'Identique 4.0

J.B. Camps – La TEI et les manuscrits

Programme

IV. La TEI et les manuscrits : La transcription des sources

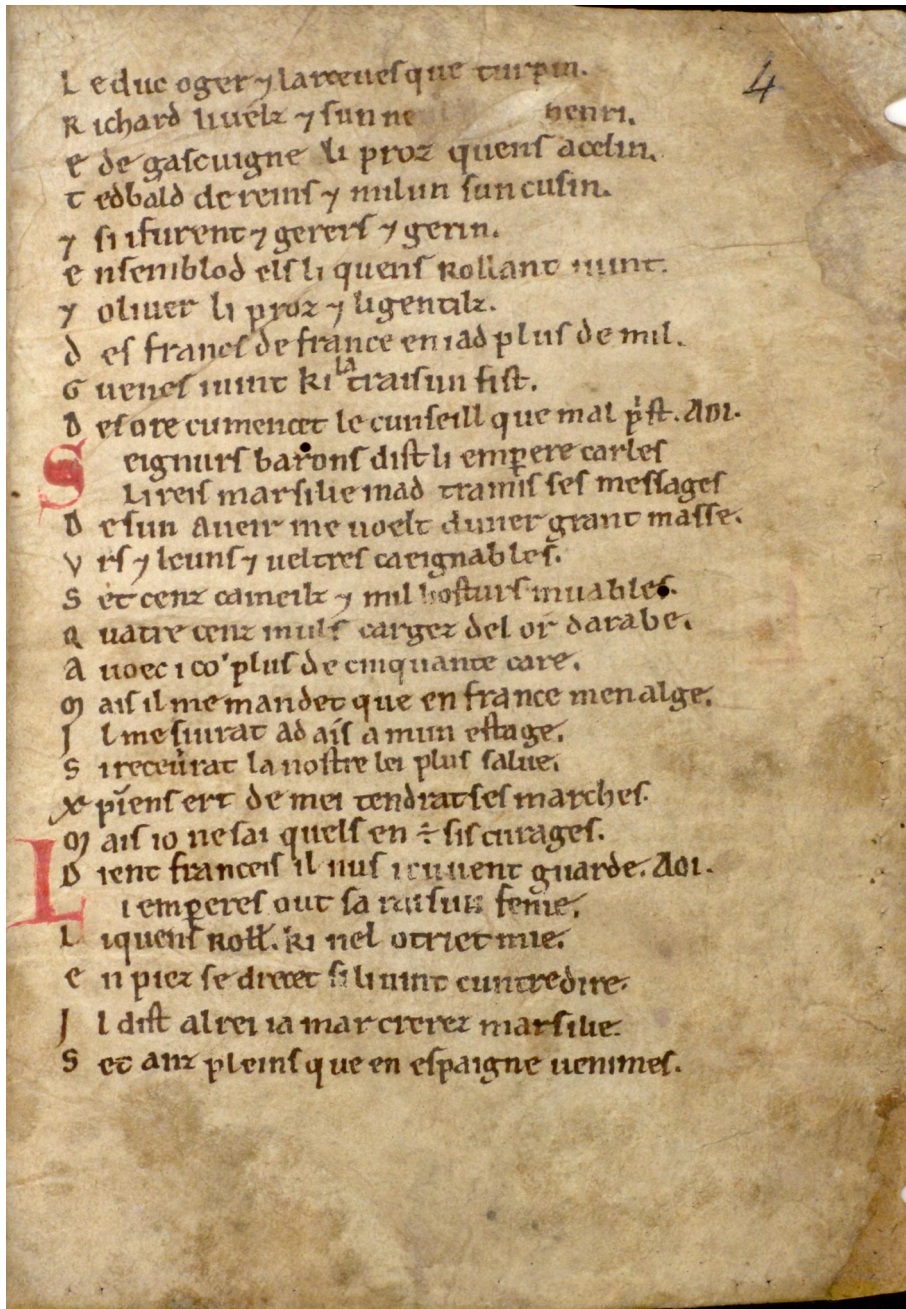
V. La TEI et les manuscrits : notices des manuscrits

VI. Édition critique et apparat

Cas d'étude et travaux pratiques :

Les manuscrits médiévaux romans : les manuscrits épiques et le *Roland* d'Oxford

Le *Roland* d'Oxford (Oxford, Bodleian Library, Digby 23)



- Plus ancien témoin conservé de chanson de geste (et de *Roland*) ;
- anglo-normand ;
- c. 1130-1140 ;
- version « archaïque » du texte (fin XIe).

Objet des T.D. :

1. édition représentant le système graphique de ce témoin (et donnant une version régularisée) ;
2. description du manuscrit.

IV. Représentation des sources primaires

Sujets : Lien facsimile/transcription ; transcriptions imitatives et éditions à visée paléographique

Guidelines, « 11 Representation of Primary Sources »

Modules transcr (*Transcription of primary sources*) et gaiji (*Characters, Glyphs, and Writing Modes*)

J.B. Camps – Mercredi 29 novembre 2017

M2 *Technologies numériques appliquées à l'histoire*
M2 *Humanités numériques*

Édition à visée paléographique

- Représentation du **système graphique** ;
- **Analyse paléographique et linguistique** traditionnelle et computationnelle ;
- Développement de méthodes d'alignement, d'identification de la mise en page, d'acquisition du texte (OCR), etc.

Voir par ex., Peter Robinson et Elizabeth Solopova, «Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue », *The Canterbury Tales Project Occasional Papers*, (1993),

<http://server30087.uk2net.com/canterburytalesproject.com/pubs/op1-transguide.pdf>

Dominique Stutzmann, « Paléographie statistique pour décrire,

identifier, dater. . . Normaliser pour coopérer et aller plus loin ? », dans *Kodikologie und Paläographie im digitalen Zeitalter*, dir. Fischer Franz, Christiane Fritz et Georg Vogeler, Norderstedt, (Schri en des Instituts r Dokumentologie und Editorik), <https://halshs.archives-ouvertes.fr/halshs-00596970/>.

Unicode et fontes pour médiévistes

MUFI (Medieval Unicode Font Initiative)

<http://www.mufi.info/>

Œuvrer pour la création de caractères Unicode répondant aux besoins des médiévistes, et élaborer une norme qui puisse être reprise par des créateurs de fontes (voir la liste, <http://www.mufi.info/fonts/>).

Quelques exemples :

- Junicode (<http://junicode.sourceforge.net/>) :

S 1 lapucele cūmande e 10l otrie / D 1ft beliffant e10 me tient pur garie

- Palemonas MUFI (<http://www.mufi.info/fonts/>) :

S 1 lapucele cūmande e 10l otrie / D 1ft beliffant e10 me tient pur garie

Et d'autres encore...

Gestion des fontes sous Linux

- via le *Visionneur de polices*
- via la ligne de commande

Installation manuelle : *dans* `/usr/share/fonts` *ou*
`~/.fonts` *ou* `~/.local/share/fonts`
`sudo fc-cache -fv`
`fc-list`

Pour Junicode, via un *package*
`sudo apt-get install fonts-junicode`

TP

- Installer Junicode ;
- Réaliser un première encodage « courant » du texte des fol. 4 et 4v de *Roland*

2.1 Correspondance fac-similé et transcription

ex. Le *Didascalicon* d'Hugues de Saint-Victor (*Thélème*, 100),
en ligne :

<http://theleme.enc.sorbonne.fr/dossiers/notice100.php>

Mise en correspondance entre facsimile et transcriptions

Nécessité d'accomplir **deux opérations** :

1. Définir et délimiter des zones de l'image, d'une part, et des zones de texte, de l'autre ;
2. Mettre en correspondance les seconds avec les premiers

Délimitation des zones du facsimile :

<facsimile/>

<facsimile/> « contient une représentation d'une source écrite quelconque sous la forme d'un ensemble d'images plutôt que sous la forme d'un texte transcrit ou encodé. »
(ou <sourceDoc/> pour une édition génétique)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" >  
  <teiHeader/>  
  <facsimile> <!-- ... --> </facsimile> <!-- Élément répétable si  
besoin -->  
  <text/>  
</TEI>
```

<facsimile/>

Éléments

<surfaceGrp/> : regroupe des éléments surface (ex., recto et verso d'un même feuillet)

<surface/> : définit une surface (ex. une page)

<zone/> : définit une zone à l'intérieur de l'élément surface parent

<graphic/> avec @url pour définir les emplacements des images

Attributs

att.coordinated (pour <surface/> et <zone/>)

@ulx (upper left x), abscisse du coin supérieur gauche (ex. 5)

@uly (upper left y), ordonnée du coin supérieur gauche (ex. 140)

@lrx (lower right x), abscisse du coin inférieur droit

@lry (lower right y), ordonnée du coin inférieur droit

ainsi que

@points, pour définir une zone non rectangulaire exprimée comme une série (séparée par des espaces) de coordonnées de points, sous la forme x,y x',y' x'',y'' ... (ex. 220,100 300,210 170,250 123,234)

NB : les coordonnées sont généralement exprimées en pixel. Elles sont facilement obtenues avec un logiciel de retouche d'image, comme **Gimp** ou de manière plus automatisée, par un algorithme d'analyse d'image

Ex. : Le *Didascalicon* d'Hugues de Saint-Victor (*Thélème*, 100)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader/>
  <facsimile>
    <surface>
      <graphic url="fax.jpg" width="693px" height="1100px"
xml:id="fax"/>
      <zone xml:id="zone__1" ulx="213" uly="3" lrx="342" lry="67"/>
      <zone xml:id="zone__2" ulx="29" uly="31" lrx="214" lry="67"/>
      <zone xml:id="zone__3" ulx="81" uly="65" lrx="335" lry="91"/>
      <zone xml:id="zone__4" ulx="80" uly="89" lrx="342" lry="117"/>
    </surface>
  </facsimile>
  <text/>
</TEI>
```

(exemple tiré de l'édition d'une page du *Didascalicon* d'Hugues de Saint-Victor (Paris, Bibliothèque Mazarine, 717, 93v° ; voir <http://theleme.enc.sorbonne.fr/dossiers/notice100.php>)

En pratique

Ouvrir avec *Gimp* le facsimile du *Didascalion* et vérifier que les coordonnées exprimées soient correctes

att.global.facs et @facs

classe d'attribut s'ajoutant aux attributs globaux, (i.e., pouvant être portés par *tous les éléments* de la TEI, dont les marqueurs <pb/>, <cb/>, <lb/>, ... mais aussi tous les autres !

att.global.facs « attributs utilisables pour les éléments correspondant à tout ou partie d'une image, parce qu'ils contiennent une représentation alternative de cette image, généralement mais pas nécessairement, une transcription »

@facs « (fac-similé) pointe directement vers une image ou vers une partie d'une image correspondant au contenu de l'élément. »

N.B. :

Il y a une certaine logique à faire porter l'attribut @facs, renvoyant à des portions du facsimile, à des éléments renvoyant à la structure physique du ms. ou à la représentation matérielle du texte, à savoir :

<lb/> (début de ligne)

<cb/> (début de colonne)

<pb/> (début de page)

<gb/> (début de cahier)

<milestone/> (début de quelque chose)

(voir *Guidelines*, « **3.10.3 Milestone Elements** »)

mais on peut également utiliser selon les besoins tout type d'élément, ou créer des segments arbitraires (avec **<seg>** par exemple)

Ex. 1 : exemple minimal, tiré des *Guidelines*

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" >
  <teiHeader>
    <!--...-->
  </teiHeader>
  <text>
    <pb facs="page1.png" />
    <!-- text contained on page 1 is encoded here -->
    <pb facs="page2.png" />
    <!-- text contained on page 2 is encoded here -->
  </text>
</TEI>
```

Ex. : Le *Didascalicon* d'Hugues de Saint-Victor (*Thélème*, 100)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader/>
  <facsimile>
    <surface>
      <graphic url="fax.jpg" width="693px" height="1100px" xml:id="fax"/>
      <zone xml:id="zone__1" ulx="213" uly="3" lrx="342" lry="67"/>
      <zone xml:id="zone__2" ulx="29" uly="31" lrx="214" lry="67"/>
      <zone xml:id="zone__3" ulx="81" uly="65" lrx="335" lry="91"/>
      <zone xml:id="zone__4" ulx="80" uly="89" lrx="342" lry="117"/>
    </surface>
  </facsimile>
  <text>
    <body>
      <div type="transcription" facs="#fax">
        <p>
          <seg facs="#zone__1">De tribus generibus lectorum. </seg>
          <seg facs="#zone__2">Satis ut puto aperte </seg>
          <seg facs="#zone__3">demonstratum e<expan>st</expan> pro vectis et
            aliquid </seg>
          <seg facs="#zone__4">amplius de se
            p<expan>ro</expan>mittentib<expan>us</expan>
            n<expan>on</expan> idem e<expan>ss</expan>e </seg>
          <!-- etc. -->
        </p>
      </div>
    </body>
  </text>
</TEI>
```

(exemple tiré de l'édition d'une page du *Didascalicon* d'Hugues de Saint-Victor (Paris, Bibliothèque Mazarine, 717, 93v° ; voir <http://theleme.enc.sorbonne.fr/dossiers/notice100.php>)

<TEI> Une autre possibilité : la mise en correspondance via linkGrp

```
<teiHeader/>
```

```
<facsimile>
```

```
<surface>
```

```
<graphic url="fax.jpg" width="693px" height="1100px" xml:id="fax"/>
```

```
<zone xml:id="zone__1" ulx="213" uly="3" lrx="342" lry="67"/>
```

```
<zone xml:id="zone__2" ulx="29" uly="31" lrx="214" lry="67"/>
```

```
</surface>
```

```
</facsimile>
```

```
<text>
```

```
<body>
```

```
<div type="transcription" facs="#fax">
```

```
<p>
```

```
<seg xml:id="seg__001">De tribus generibus lectorum. </seg>
```

```
<seg xml:id="seg__002">Satis ut puto aperte </seg>
```

```
</p>
```

```
</div>
```

```
<linkGrp>
```

```
<link targets="#zone__1 #seg__001"/>
```

```
<link targets="#zone__2 #seg__002"/>
```

```
</linkGrp>
```

```
</body>
```

```
</text>
```

```
</TEI>
```

T.P. Roland

1. Créer un élément <facsimile/> pour le fol. du ms. de la *Chanson de Roland*, et y définir des groupes de surface, des surfaces, et des zones ;
2. Mettre en correspondance les sections de la transcription avec les zones définies. Raccrocher de préférence ces indications à des éléments déjà présents, sans créer de segments arbitraires.

Alignement texte/image et ROC du manuscrit

L'alignement texte et image peut tant servir à l'établissement d'édition à visée paléographique ou linguistique, que permettre la fourniture de données utiles pour perfectionner la reconnaissance optique des caractères (anglais : *Optical character recognition*) pour le manuscrit

- Projet ORIFLAMMS, voir <https://oriflamms.hypotheses.org/> et « Oriflamms Alignment Software », cf. A. Lavrentiev, Y. Leydier, D. Stutzmann, « Specifying a TEI-XML Based Format for Aligning Text to Image at Character Level », Symposium on Cultural Heritage Markup, 2015, en ligne : <http://www.balisage.net/Proceedings/vol16/print/Lavrentiev01/BalisageVol16-Lavrentiev01.html>
- Transkribus, <https://transkribus.eu/Transkribus/>.

Détection des lignes

Ces outils intègrent généralement des algorithmes d'analyse de la mise en page et de détection des zones de texte. Par exemple, Ocropy (<https://github.com/tmbdev/ocropy>)

Étapes principales

- 1 Acquisition des images (de préférence tif avec une résolution d'au moins 300 DPI) ;
- 2 Binarisation ;
- 3 détection des colonnes et des lignes ;
- 4 acquisition du texte
- 5 si erreurs, corriger et réentraîner le modèle
- ...
- 6 extraction des données textuelles, éventuellement de l'alignement avec l'image

2.2 TEI et éditions allographiques

Numérisation du ms.

Oxford, Bodleian Library, Digby 23

<http://image.ox.ac.uk/show?collection=bodleian&manuscript=msdigby23b>

4
Le due oger 7 larceues que curpin.
Richard liuelz 7 sun ne henri.
E de gascuigne li proz quent acelin.
E edbald de rems 7 milun sun cusin.
7 si ifurent 7 gerent 7 gerin.
E nsembloz els li quent rollant nunt.
7 oliuer li proz 7 ligentelz.
Des francz de france en iad plus de mil.
Guenel nunt ki traifun fist.
Desore cumencet le conseil que mal pft. doi.
Seignurs barons dist li emperer carles
Li reis marsilie mad tramis ses messages
Desun auer me uoelt duner grant masse.
Yris 7 leuns 7 ueltes caignables.
Set cens cameils 7 mil hosturs muables.
Quatre cens mult charges del or darabe.
A uoec i co' plus de cinquance care.
Oz ail il me mandet que en france menalge.
Il mesurat ad ail a mun estage.
Sireceirat la nostre lei plus salue.
X pieus ert de mei tendrat ses marches.
Oz ail io ne sai quell en est li curages.
Lient francz il nuf reuuent garde. doi.
Li emperer out sa raisus feme.
Li quent noth. ki nel ocriet mie.
En piez se drecet si liunt cuncredre.
Il dist al rei ia mar crevez marsilie.
Sees aux pleins que en espaigne uenmes.

Li nos cuqst 7 noptes 7 comibles.
Puis aqualoz rne 7 la terre de pme.
7 balasqued 7 cuele 7 sezilie.
Li reis marsilie ifist mult que traire.
Deses parens iat quinze.
E hancuns portout une branchedolue.
Huncerent nos oez paroles meisme.
A uoz francz un conseil enprentes.
Loerent uos aloz de legerie.
Douz de uoz cuntes al paien tramesistes.
Lun fut basan 7 li autres basilies.
Les chef en pft espuis de suz bataille.
Faites la guer cu uos lauez enpse.
En sarraguce menez nostre ost bame.
Loiez de sege a tute ure ue.
Suengez cels que li reis fist ocire. doi.
Ie mpe en tuit sun chef enbrunc.
Sidiust sabarbe a fatad sun gernun.
He ben ne mal ne respunt sun neuid.
Francz se tarsent ne mais que guenelun.
En piez se drecet siunt deuant carlun.
Oult fierent cumcet sa raisun.
E dist al rei ia mar crevez brieun.
Ke mei ne altre se de uostre pd nun.
Quant co nos mandet li reis marsilun
Ql denendrat iointes ses mains tuf hom.
E tute espaigne tendrat par nostre dun.

Modélisation

Quels éléments veut-on représenter ?

Modélisation

Quels éléments veut-on représenter ?

Systeme graphique : allographes, abréviations, segmentation, ponctuation ancienne...

Interventions scribales : ajouts, suppressions, corrections, ...

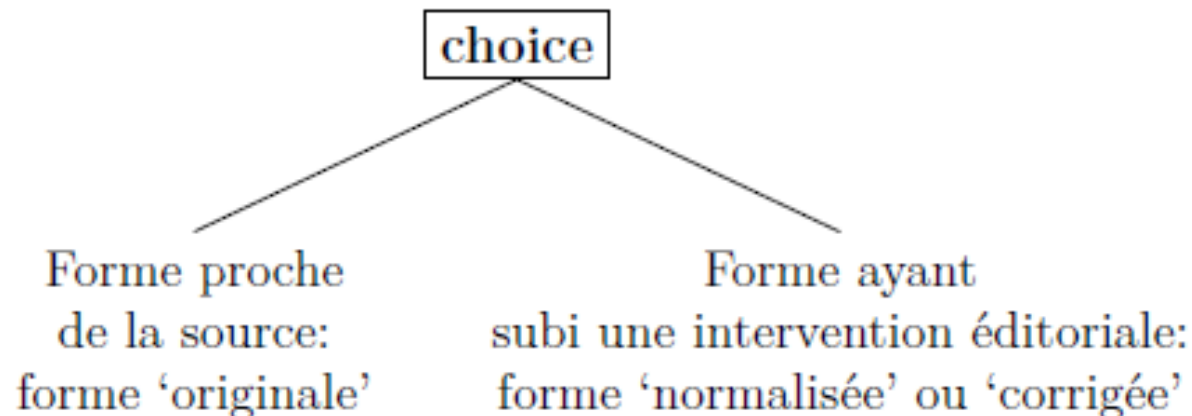
Interventions éditoriales : normalisations, corrections, difficultés de transcription, ...

2.2.1 une représentation double avec <choice>

Permet la représentation simultanée de deux états du texte

Peut contenir tous les éléments du **model.choicePart** :

sic/corr ; reg/orig ; unclear ; abbr/expan ; ex, am et seg



Abréviations

« 17 oc^{bre} 1822 »

Solution utilisable avec TEI-Lite

17

<choice>

<abbr>oc<hi rend="sup">bre</hi></abbr>

<expn>octobre</expn>

</choice>

1822

Utilisation de ex pour marquer les lettres restituées

17

<choice>

<abbr>oc<hi rend="sup">bre</hi></abbr>

<expn>oc<ex>to</ex>bre</expn>

</choice>

1822

Solution sans choice (plus limitée) :

17 oc<ex>to</ex>bre 1822

NB : am peut être utilisé si des lettres figurent dans l'abréviation et pas dans sa résolution. Ex. ss.
pour sancti

Traitement des allographes

```
<choice>  
  <reg>s</reg>  
  <orig>ʃ</orig>  
</choice>
```

```
<choice>  
  <reg>s</reg>  
  <orig>&#383;</orig>  
</choice>
```

Quelques limites de cette première solution

- les *Guidelines* tendent à recommander un emploi de `expan`, `reg`, ..., au niveau du mot (`expan` « should usually be a complete word or phrase ») ;
- détournement du sens de caractères Unicode
par. ex., si l'on utilise U+0303 COMBINING TILDE (~) pour la tilde abrégative médiévale, alors que son sens est « IPA: nasalization / Vietnamese tone mark »

2.2.2 Définir des caractères (module gaiji)

En toute bonne rigueur, lorsque l'on utilise des caractères qui n'ont pas de correspondant exact dans Unicode, que l'on détourne des glyphes de fontes Unicode*, ou que l'on veut décrire spécifiquement des variantes de caractères Unicode existants, il importe de clarifier cela en utilisant des éléments du module gaiji, cf. *Guidelines*, « **5 Characters, Glyphs, and Writing Modes** »), pour soit annoter un caractère existant (précision à apporter), soit créer un nouveau caractère.

⚠ Ne pas confondre *caractère* (symbole abstrait) et *glyphe* (une représentation physique donnée, une réalisation d'un caractère)

* (par exemple, le caractère U+0303 COMBINING TILDE, qui représente la nasalisation dans l'API, pour représenter la tilde abrégative médiévale).

Définir des caractères (module `gaiji`)

Deux éléments :

Dans le `teiHeader`

charDecl pour définir ou annoter des caractères dans le `teiHeader/encodingDesc` (`model.encodingDescPart`), à l'aide d'éléments

char : définit un nouveau caractère

glyph : précise une réalisation (un glyphe) d'un caractère existant

Dans le corps du document

g (et `@ref`) pour faire référence à un glyphe dans le corps du document (`model.gLike`)

char / glyph

- **charName** ou **glyphName** pour préciser le nom d'un caractère ou d'un glyphe ;
- **gloss** et **desc**
- **charProp** (character property) permet de décrire des propriétés du caractère (paires nom/valeur), avec `unicodeName/localName` et `value` ;
- **mapping** (character mapping) « contient un ou plusieurs caractères reliés par certains aspects (spécifiés par l'attribut `type`) au glyphe ou au caractère défini dans l'élément parent »
- **figure**
- **note**

Exemple de définition de caractère (*Guidelines*)

<charDecl>

<desc>Description des caractères et glyphes :

Exemples tirés des Guidelines</desc>

<char xml:id="aenl">

<charName>LATIN LETTER ENLARGED

SMALL A</charName>

<charProp>

<localName>entity</localName>

<value>aenl</value>

</charProp>

<mapping type="standard">a</mapping>

</char>

<!-- Autres définitions -->

</charDecl>

Exemple d'annotation (*Guidelines*)

<charDecl>

<desc>Description des caractères et glyphes : Exemples
tirés des Guidelines</desc>

<!-- ... -->

<glyph xml:id="r1">

<glyphName>LATIN SMALL LETTER R WITH ONE
FUNNY STROKE</glyphName>

<charProp>

<localName>entity</localName>

<value>r1</value>

</charProp>

<figure>

<graphic url="r1img.png"/>

</figure>

</glyph>

<!-- ... -->

</charDecl>

Répertoires de définitions de caractères

James Cummings, « The **ENRICH Gaiji Bank** », *Manuscriptorium : Building Virtual Research Environment for the Sphere of Historical Resources*, en ligne :

http://v2.manuscriptorium.com/apps/gbank/gbank_table.php

Définir des caractères (module `gaiji`)

Utiliser les caractères dans le corps du document :

`<p>`Attention, ce `<g ref="#r1">r</g>` a une forme bizarre. Et ce `<g ref="#aenl">a</g>` n'existait pas dans Unicode. `</p>`

N.B. : la TEI définit un élément `<c>` (caractère), de la même série que `<w>` (*word*) ou `<s>` (*sentence*), qui peut aussi être utilisé pour spécifier le rendu d'un caractère précis.

`<c rend="lettrine">A</c>`ttention

2.2.3 utilisation d'une DTD et d'entités

La déclaration de DTD

DTD externe

```
<!DOCTYPE nomÉlémentRacine SYSTEM "madtd.dtd">
```

DTD interne

```
<!DOCTYPE nomÉlémentRacine [  
  <!ELEMENT nomÉlémentRacine (#PCDATA)>  
  Définition des autres éléments...  
>
```

DTD externe avec sous-ensembles internes

```
<!DOCTYPE document SYSTEM "madtd.dtd" [  
  <!ELEMENT elementNonDecritDansmadtd (#PCDATA)>  
>
```


Rappel de cours : les entités

```
<!ENTITY entité "contenu">
```

```
<!ENTITY ENC "École nationale des chartes">
```

&ENC; sera développé par le parseur en « École nationale des chartes »

Entités

Peuvent contenir des éléments TEI :

```
<!DOCTYPE TEI [  
  <!ENTITY s-long  
'<choice><reg>s</reg><orig>&#383;</orig  
  ></choice>' >  
  ]>
```

Des difficultés de modélisation qui subsistent

- Problématique de la définition des variantes de forme ;
- Portée abrégative et son niveau.

Un exemple d'encodage du début du fragment de Mende

(Voir les fichiers)

T.P.

Mise en place

6) Ajouter une déclaration DOCTYPE, interne ou externe, au document

Définition des entités et transcription

- 7) Créer les entités nécessaires à la transcription du fragment (allographes, signes abrégatifs,...), et les utiliser où elles sont nécessaires ;
- 8) Utiliser, là où ils sont nécessaires les éléments pour représenter les lacunes matérielles, les difficultés de lecture, les notes, la séparation des mots...

Bibliographie

En complément des références données au cours de la présentation.

Texte de référence

TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>>, particulièrement
« 10 Manuscript Description »
« 11. Representation of Primary Sources »

Éditions à visée paléographique et analyse statistique

A. Lavrentiev, Y. Leydier, D. Stutzmann, « Specifying a TEI-XML Based Format for Aligning Text to Image at Character Level », *Symposium on Cultural Heritage Markup*, 2015, en ligne :

<http://www.balisage.net/Proceedings/vol16/print/Lavrentiev01/BalisageVol16-Lavrentiev01.html>

Dominique Stutzmann, « Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin ? », dans *Kodikologie und Paläographie im digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2*, dir. Franz Fischer *et al.*, 2010, p. 247-277, en ligne :

http://kups.ub.uni-koeln.de/4353/1/15__stutzmann.pdf.

La TEI et les manuscrits

V. La description des manuscrits

Sujet : Notices de mss en TEI

Guidelines, « 10 Manuscript Description »

Module msdescription

J.B. Camps – Mercredi 6 décembre 2017

M2 Technologies numériques appliquées à l'histoire

M2 Humanités numériques

2.1 Bref historique et enjeux de la description des mss et sources anciennes en TEI

1999-2001 : MASTER

1999-2001 : projet européen MASTER (*Manuscript Access through Standards for Electronic Records*)

<http://master.dmu.ac.uk/>

=> ajouts d'éléments à la TEI (`msDescription`) pour décrire les mss,

- *Reference Manual for the MASTER Document Type Definition :*

Discussion Draft, dir. Lou Burnard pour le groupe de travail MASTER, en ligne :

http://www.tei-c.org/About/Archive_new/Master/Reference/oldindex.html

- Présentation du projet à son origine : Lou Burnard, Muriel Gougerot, Elizabeth Lalou, et Peter Robinson, *Towards a European Standard for Manuscript Description: the MASTER project*, 1999, en ligne :

http://www.tei-c.org/About/Archive_new/Master/Hermes/hermes.pdf

Objectifs de *MASTER*

Objectives

1. To establish a **single interchange standard for both first-level (abbreviated) and full descriptions of manuscripts**, through two phases of development
2. To **test the standard**, refining it as needed, on descriptions of at least five thousand manuscripts, in different collections across Europe, by the end of the project
3. To **document fully the standard**, which will be developed in partnership with the Text Encoding Initiative workgroup on manuscript description
4. To **develop software systems** to permit libraries with relatively few resources to make simple descriptive records, and provide means for elaboration of these into fully-detailed records
5. To **publish at least five thousand manuscript descriptions** on the Internet, with specialized search software and linked to some manuscript images and full text manuscript transcripts, in a prototype on-line manuscript union catalogue and advanced manuscript publication system, by the end of the project
6. Through 5: to bring the manuscripts to many more people
(source : <http://master.dmu.ac.uk/>)

ENRICH

2007-2009 : lancement du projet ENRICH
(European Networking Resources and Information
concerning Cultural Heritage)

<http://enrich.manuscriptorium.com/> et de la
plateforme Manuscriptorium
www.manuscriptorium.com/

Production d'une personnalisation de la TEI, autour du module
msDescription :

The ENRICH Schema : A Reference Guide, dir. Lou Burnard, Oxford,
2008, en ligne :

http://projects.oucs.ox.ac.uk/ENRICH/Deliverables/referenceManual_en.html

Ressources en lien avec ENRICH :

<http://projects.oucs.ox.ac.uk/ENRICH/>

Différences MASTER / ENRICH

« **Earlier versions of Manuscriptorium used schemas based on MASTER [...].** These specifications both defined comparatively unconstrained XML formats, which permitted a very wide range of possibilities and did not attempt to constrain (for example) values to any predefined set of values. While appropriate for an interchange format, this approach has some drawbacks:

- **there may be wide variation in approaches taken to represent essentially the same phenomenon;**
- **the format appears over complex to novice users,** who will only ever want to use a very small subset of the possible tags;
- **developing software (e.g. stylesheets) for the format becomes unnecessarily complex,** since every possibility must be allowed for even though it is unlikely to appear;
- **accurate searching of the data may be needlessly complicated** by the large number of ways of representing e.g. attribute values such as dates.
- **In the ENRICH schema, we have tried to address these concerns by reducing the number of choices and constraining the possible values for several attributes.** Nevertheless, the resulting schema remains **fully TEI Conformant:** we are only defining a **subset;** all constraints introduced have the full consent of all partners in the project. »

(source : *The ENRICH Schema : A Reference Guide*, dir. Lou Burnard)

Le *Special Interest Group* sur les manuscrits (TEI MS SIG)

Voir : <http://wiki.tei-c.org/index.php/SIG:MSS>

Mis en place par Elena Pierazzo, Susan Schreibman, et Edward Vanhoutte. Actuellement
mené par Gerrit Brüning and Dot Porter.

Ses activités récentes

- recommandations pour les éditions « génétiques » (fait)
- révision de l'apparat critique ([Critical Apparatus Workgroup](#))
- révision des éléments pour la collation des cahiers des mss

Les révisions possibles du module msDescription

« investigate on was whether the module msDesc could be used for non-medieval non-codex manuscript or non-manuscript material or has to be extended. Possible materials include (but are not limited to)

- printed material
- text-bearing-object made of something other than parchment (tombstone, statue, bell, curse tablet) »

(source : http://wiki.tei-c.org/index.php/MsDesc_enhancement)

Descriptions de manuscrits et incunables en France

2002 : lancement de l'informatisation du Catalogue général des manuscrits des bibliothèques publiques de France (CGM) ; EAD est retenu. La même année EAD est également retenu par le Dép. des mss de la BnF.

2010 : publication par l'Agence Française de Normalisation, de la norme DeMArch (Description des Manuscrits et fonds d'Archives modernes et contemporains en bibliothèque)

2012 : début de l'informatisation des *Catalogues régionaux des incunables des bibliothèques publiques de France*, dans un format MARC-TEI développé pour l'occasion (descr. en ligne sur le site des BVH, <http://www.bvh.univ-tours.fr/incunables.asp>).

2017 : retour à un format entièrement Unimarc, pour l'intégration au CCFr, cf. Fatima Elamri Crispyn, *Intégrer les données rétro converties des catalogues régionaux d'incunables dans le Catalogue Collectif de France : problématiques scientifiques, méthodologie, modèles de conversion de données*, mém. master TNAH, 2017.

Néanmoins, existence de projets ou d'institutions faisant emploi de la TEI : utilisation par l'Institut de recherche et d'histoire des textes (IRHT), cf. Laura Lebarbey, *Reconstituer et étudier une collection ancienne au-delà de la pluralité des bases et des formats*, mém. master TNAH, 2015 ;

projet de la B.M. de Grenoble.

Références

Florent Palluault, « Informatiser des descriptions complexes : l'utilisation de l'EAD et de la TEI pour les manuscrits et les livres anciens en France », IFLA 2012 (Helsinki), en ligne < <http://conference.ifla.org/past/2012/212-palluault-fr.pdf> >.

Lydie Ducolomb, *Le catalogage des manuscrits médiévaux en France : état des lieux et perspective*, dir. F. Palluault, mémoire de fin d'étude du diplôme de conservateur, promotion DCB19, en ligne < <http://www.enssib.fr/bibliotheque-numerique/notices/49073-le-catalogage-des-manuscrits-medievaux-en-france-etat-des-lieux-et-perspectives> >.

EAD en bibliothèque : Guide des bonnes pratiques, en ligne : <http://bonnespratiques-ead.net/>.

Laura Lebarbey, *Reconstituer et étudier une collection ancienne au-delà de la pluralité des bases et des formats*, mémoire de master « Technologies numériques appliquées à l'histoire », Paris, 2015.

2.2 Modules et éléments pour l'élaboration de notices et la description des sources

Guidelines, « 10 Manuscript Description »
Module msdescription

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

Exemples et cas concrets

Oxford, Bodleian Library, Digby 23

<http://image.ox.ac.uk/show?collection=bodleian&manuscript=msdigby23b>

BnF, nouv. acq. fr. 5094

<http://gallica.bnf.fr/ark:/12148/btv1b6000371s/f19.image>

Cologne-Genève, Fondation Martin Bodmer, Cod.
Bodmer 168

<http://www.e-codices.unifr.ch/en/description/fmb/cb-0168/>

Le ms. KB, Tilander fr. 6
J.B. Camps – La TEI et les manuscrits

<msDesc>

<msDesc> (model.biblLike) : racine de la description du ms.
(dans sourceDesc ou ailleurs)

- <msIdentifier> éléments d'identification du ms., permettant d'en exprimer la cote ; seul élément obligatoire
- <head> titre
- <msContents> description du contenu du ms.
- <physDesc> description physique du ms.
- <history> histoire du ms.
- <additional> informations complémentaires

- <msPart> subdivision du ms. pouvant contenir sa description complète ; à utiliser surtout pour les **recueils factices**, composés de plusieurs unités codicologiques distinctes.
- <msFrag> (nouveau) sert à décrire des fragments, conservés séparément, d'un même manuscrit

Points de repères dans le ms.

locus définit un emplacement dans le ms., généralement sous la forme d'un feuillet ou d'une série de feuillets (@scheme, @from, @to)

locusGrp permet de décrire une série d'emplacements (locus)

Quantités, dimensions et support

measure décrit la mesure d'une quantité d'un objet

dimensions description de dimensions (@type)

objectType description d'un type d'objet

material description du matériau de quelque chose

watermark description d'un filigrane

Date et lieu d'origine

origDate date pour le ms. ou une partie du ms.

origPlace lieu d'origine pour le ms. ou une partie du ms.

Structuration physique du ms.

secFol second folio

catchwords description des réclames

signatures description des signatures de cahier

Marques et provenance

heraldry description d'un élément de décor héraldique

stamp : estampille

<msIdentifier>

Une cote de ms. :

Cote abrégée : Paris, BnF, fr. 854

Version plus complète : France, Paris, Bibliothèque nationale de France, Département des manuscrits, fonds français 854 (Chansonnier occitan *I*) *olim* Regius 7225, *olim* Rigault II 1880, *olim* Dupuy I 1959

Un exemple utilisant une descr. structurée

```
<msIdentifier>  
  <country>France</country>  
  <settlement>Paris</settlement>  
  <institution>Bibliothèque nationale de France</institution>  
  <repository>Département des manuscrits</repository>  
  <collection>Fonds français</collection>  
  <idno>854</idno>  
  <msName>Chansonnier occitan I</msName>  
  <altIdentifier type="ancienne_cote">  
    <institution>Bibliothèque royale</institution>  
    <collection>Regius</collection>  
    <idno>7225</idno>  
  </altIdentifier>  
  <!-- etc. -->  
</msIdentifier>
```

Quelques exemples suivant le schéma ENRICH

- **Cod. Bodm. 168 (Waldef, Guy de Warwick, Otinel)**
<http://www.e-codices.unifr.ch/fr/description/cb/0168>

- **Stockholm, KB A 148 (« Codex Gigas »)**
http://www.manuscriptorium.com/apps/main/en/index.php?request=request_document&docId=set20070920__12__84&mode=&client

<head>

Le titre du ms. (tel qu'on le donne dans la notice, pas forcément le nom d'usage)

Ex. 1

```
<head>Petrus Comestor, Historia scholastica</head>
```

Ex. 2

```
<head>
```

```
  <title>Waldef</title>
```

```
  <origDate notBefore="1270" notAfter="1330">Fin du XIIIe  
siècle-début du XIVe siècle</origDate>
```

```
</head>
```

```
<head>
```

```
  <title>Gui de Warewic</title>
```

```
</head>
```

```
<head>
```

```
  <title>Otinel</title>
```

```
</head>
```

<msContents>

Description du contenu du ms. (entités intellectuelles) sous forme de paragraphes **ou** d'éléments `<msItem>/<msItemStruct>`, subdivisés en

- `locus` (ou `locusGrp`) pour les feuillets occupés par le texte
- `author` ou `respStmt`
- `title`
- `rubric` (pour la rubrication)
- `incipit` / `quote` / `explicit`
- `finalRubric` rubrication finale
- `colophon`
- `decoNote` note sur la décoration
- `listBibl` / `bibl` références bibliographiques (par exemple les éditions du texte)
- `filiation` éléments sur la place de cette copie du texte dans la tradition (copie, modèle, ...)
- `note`
- `textLang` langue(s) du texte

Lier les informations à des référentiels

att.canonical

@key provides an externally-defined means of identifying the entity (or entities) being named, using a coded value of some kind.

@ref (reference) provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URIs. «

Ex.

```
<title key="Otinel" ref="http://viaf.org/viaf/199768389">Chanson  
d'Otinel</title>
```

Ex.

```
<author key="Brézé, Jacques de, 14..-1494"  
ref="http://viaf.org/viaf/24739331">Jacques de Brézé</author>
```

<physDesc>

objectDesc description de l'objet livre, du **codex** a proprement parler et de sa fabrication : support, cahiers, préparation de la page (réglure, piquêre, etc.)

Copie et décoration : écritures, les différentes mains, la décoration, les annotations, gloses et traces d'utilisation

Reliure : la description de la **reliure**, des fermoirs, signets, seaux, etc.

physDesc/objectDesc

supportDesc description du support et de la constitution des cahiers

- **support** la nature du support, pouvant inclure des éléments généraux tels que **material**, **watermark**, **dimensions** ,
- **extent** nombre de feuillets
- **foliation** système de foliotation
- **collation** collation des cahiers (nombre de cahiers, nombre de bifeuillets par cahier, ...) ; peut faire appel à des éléments comme **locus**, **formula**, **catchwords** (réclames), **signatures** (signatures de cahier), etc.
- **condition** condition matérielle du ms.

layoutDesc pour la description de la mise en page, comportant un ou plusieurs éléments **layout**

```

<physDesc>
  <objectDesc>
    <supportDesc>
      <support><material>Papier</material>, <dimensions unit="mm">
        <height>185</height>
        <width>135</width>
      </dimensions>,
      <watermark>filigrane figurant la lettre P
      fourchue ?</watermark>
    </supportDesc>
    <extent>
      <measure type="composition" unit="feuillet">16</measure>
    </extent>
    <foliation>1 fol. de garde + foliotation moderne 1-16</foliation>
    <collation>
      <formula>1:16</formula>
      <catchwords>Réclames au fol. <locus>x</locus></catchwords>
    </collation>
    <condition>vraisemblablement endommagé (humidité) au cours de la seconde
    moitié du XIXe ou première du XXe ; lourdement restauré.</condition>
  </objectDesc>
  <!-- suite -->
</physDesc>

```

Formules de collation

Notation de Chroust (cf. DFG, *Richtlinien Handschriftenkatalogisierung*, voir http://www.manuscripta-mediaevalia.de/hs/katalogseiten/HSKRICH_a10_jpg.htm)

$$I^2 + (V-1)^{11} + 21 IV^{169} + (III-1)^{184}$$

```
<formula corresp="vol.1" notation="Chroust"> I<hi rend="sup">2</hi> +  
(V-1)<hi rend="sup">11</hi> + 21 IV<hi rend="sup">169</hi> +  
(III-1)<hi rend="sup">184</hi></formula>
```

Tableau

```
<row>  
  <cell>1</cell>  
  <cell>  
    <locus from="1r" to="8v">1-8</locus>  
  </cell>  
  <cell>4-4</cell>  
  <cell> </cell>  
  <cell>réclame</cell>  
  <cell>Les feuillets ont été cousus ainsi:<lb/> 3-6<lb/> 4—  
5<lb/> 2—7<lb/>1—8</cell>  
</row>
```

```
<layoutDesc>
```

```
<layout columns="2" writtenLines="42">
```

```
<p>Justification:
```

```
<dimensions unit="mm">
```

```
<height>220</height>
```

```
<width>130</width>
```

```
</dimensions>.
```

```
Texte sur deux colonnes; sauf  
exception, 42 lignes par colonne. Réglure à l'encre.</p>
```

```
</layout>
```

```
</layoutDesc>
```

Formules de réglure

Systeme Leroy

cf. D. Muzerelle, « Présentation critique du « système Leroy » pour la description des schémas de réglure », en ligne :
<http://www.palaeographia.org/muzerelle/codeleroy.htm>

Systeme de Muzerelle

Denis Muzerelle, « Analyse des réglures
Diaporama didactique », en ligne :

<http://www.palaeographia.org/muzerelle/reglure.htm>

-- , « MASTARA – Traçage automatique
de schémas de réglure », en ligne :

<http://www.palaeographia.org/muzerelle/mastara.htm>

physDesc : 2. description de l'écriture, des mains, etc.

- **handDesc** description des mains, comme un paragraphe, ou une série de **handNote** pour chacune des mains ; la classe d'attributs `att.handFeatures` (`@scribe`, `@scribeRef`, `@script`, `@scriptRef`, `@medium`, `@scope`) permet d'en préciser certaines caractéristiques
- **scriptDesc** description des écritures, § ou **scriptNote** (ou leurs équivalents **typeDesc** et **typeNote** pour les fontes des imprimés)
- **decoDesc** description de la décoration, § ou **decoNote**
- **musicNotation** description des notations musicales, si elles sont présentes
- **additions** description des ajouts principaux (gloses, *marginalia*, etc.)

Ex. : descr. des mains et de l'écriture

```
<handDesc hands="2">  
  <handNote xml:id="main1" scope="major"> <p><locus from="1" to="14">fol. 1-14</locus>  
Première main, dans une écriture gothique textualis (...)</p></handNote>  
  <handNote xml:id="main2" scope="minor"> <p><locus from="15" to="18">fol. 14-  
15</locus> Seconde main, plus tardive, dans une cursiva gothique (...)</p></handNote>  
</handDesc>  
<scriptDesc>  
  <scriptNote scribeRef="#main1"><p>Écriture gothique textualis utilisée par la première  
main (...)</p></scriptNote>  
  <scriptNote scribeRef="#main2"><p>Écriture gothique cursiva utilisée par la seconde  
main (...)</p></scriptNote>  
</scriptDesc>
```

L'attribut @hand peut-être utilisé dans le document pour y faire référence ; l'élément
<handshift> pour signaler le changement de main

```
<add hand="#main2" cert="high" resp="#JBC"> <!--addition de la main 2--></add>
```

```
<handShift new="#main2"/> <!--Passage de la main 1 à la main 2 dans le texte-->
```

physDesc : 3. la reliure, les sceaux, ...

- **bindingDesc** description de la ou des reliures du ms. (reliure actuelle et éventuellement reliures anciennes conservées à part ou perdues), chacune dans un élément **binding**
- **sealDesc** description du ou des sceaux, dans des éléments **seal** (éventuellement complétés d'éléments **decoNote**).
- **accMat** (« accompanying material ») éléments extérieurs au ms. en lui-même qui ne rentrent pas dans les cases précédentes (ex. fragment utilisé pour la reliure, etc.)

<bindingDesc>

<binding notBefore="1701-01-01" notAfter="1800-12-31">

<p>Reliure du XVIIIe siècle, en

<material>maroquin rouge</material> ,

portant des armes, <heraldry>(descr. des
armes)</heraldry> <!--etc. --> </p>

</binding>

</bindingDesc>

Histoire du ms.

history contenant

- **origin** informations sur l'origine du ms.
- **provenance** la ou les provenances du ms et ses possesseurs anciens
- **acquisition** acquisition du ms. (c'est-à-dire son entrée dans la collection envisagée).

Mise en pratique

Tenter de structurer la notice du ms. Oxford,
Bodleian Library, Digby 23
et pourquoi pas des
BnF, NAF 5094 ou KB, Til. fr. 6

Bibliographie : TEI et notices de mss

En complément des références données au cours de la présentation.

Texte de référence

TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*,
<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>>, particulièrement
« 10 Manuscript Description »
« 11. Representation of Primary Sources »

The ENRICH Schema — A Reference Guide, dir. Lou Burnard, Oxford, 2008, en ligne
<http://projects.oucs.ox.ac.uk/ENRICH/Deliverables/referenceManual_en.html>.
Documentation particulière du schéma Enrich.

Sur la description des incunables, voir :

les actes de la journée *Signalement et numérisation des incunables*, Centre d'études
supérieures de la Renaissance, Tours, 18 novembre 2010, en ligne :

http://www.patrimoineecrit.culture.gouv.fr/catalogues/outils-referance/je-incunable_s.php