

# Atelier « Introduction à l'édition électronique savante »

Pourquoi « savante » et pas « critique » ? Parce que le critère d'édition critique ou non peut être sujet à débat selon les écoles, et peut-être aussi pour correspondre à ce qui est appelé en anglais « Digital Scholarly Editing », qui est plus large que « Critical editing ».

(D) Sur cette question, je mets ici quelques références bibliographiques récentes qui ne sont pas exhaustives mais qui font un peu le point sur le débat autour de la théorie de l'édition critique électronique. Pour que vous puissiez récupérer toutes les ressources, je mettrai cette présentation à disposition tout comme ce que j'ai raconté mardi après-midi.

## (D) L'édition électronique, qu'est-ce ?

(D) déf. de l'édition critique électronique de P. Robinson dans son article de 2002, 6 critères :

- comporte une analyse historique du ou des documents
- émet des hypothèses au sujet de la création et de la variation du texte
- apporte description et classification des matériaux
- peut offrir un texte édité, c'est-à-dire reconstitué, entre autres
- offre des possibilités aux lecteurs de concevoir leurs propres hypothèses et choisir leurs modes de lecture
- tout cela étant publié d'une manière enrichie

De ces critères, les quatre premiers, voire le sixième, s'appliquent aussi aux éditions critiques non numériques, ce qui pose un problème de définition du périmètre réel de l'édition critique électronique.

(D) Une autre déf. est celle de Patrick Sahle dans le volume collectif **Digital Scholarly Editing de 2016** :

- Une édition scientifique est la représentation critique d'un document ou d'un texte historique
- une édition numérisée (sous-entendu une édition critique papier qui aurait été numérisée) n'est pas une édition numérique (sous-entendu nativement numérique)
- une édition numérique ne peut pas être imprimée sans perte significative de contenu et de fonctionnalités.

Cette définition par la négative permet de faire ressortir la particularité de l'édition numérique par rapport à une édition papier.

(D) P. Sahle en conclut que les éditions scientifiques numériques sont (ou devraient être) des éditions qui sont guidées par un « paradigme digital » en ce qui concerne leurs fondements théoriques, leur méthodologie et leur mise en pratique, c'est-à-dire qu'elle devraient être pensées dès le départ pour le support numérique.

## (D) L'édition électronique, pourquoi, pour qui ?

### Les avantages

- (D) dépasser les limites physiques de la page
- permettre de fournir tout le matériel qui a servi à l'éditeur à préparer son édition, ce qui satisfait à la fois les critères de transparence, d'exhaustivité et de précision
- permettre d'atteindre un public plus varié : le degré d'extrême spécialisation des appareils critiques les rend lisibles seulement par des chercheurs en philologie, comme le souligne Elena Pierazzo dans le même volume *Digital scholarly editing*

### La question du public visé

- (D) un autre avantage de l'édition électronique est donc la possibilité d'adapter le contenu au profil du lecteur ou plutôt au type de lecture désiré, qui est laissé à l'appréciation de l'utilisateur
- il faut donc prévoir des parcours de lecture et pour ce faire, encoder les éléments à ne pas montrer pour pouvoir les disposer uniquement à la demande

### Les risques

- (D) noyer le lecteur sous un flot d'information sans proposer un texte unique lisible, et donc, ne s'adresser qu'à un type d'utilisateur, le chercheur, qui pourra éventuellement trouver son compte dans ce qui pourrait être décrit comme un « bruit »
- ceci oblige à poser la question suivante : une édition électronique peut-elle être encore qualifiée de critique ?

Oui, si l'on prend en compte une définition ouverte du terme « critique ».

=> le terme s'applique donc ici à toute édition qui fournit un enrichissement par rapport au document source, enrichissement obtenu par l'application d'une analyse réflexive dans différents domaines :

- correction des erreurs de copie
- typologie et classement des variantes
- application de règles dans la transcription
- indexation, identification des sources
- rédaction d'une introduction et d'une description des manuscrits
- choix de publication

Notons qu'on peut tout à fait réaliser une édition critique électronique respectant les normes traditionnelles pour cette discipline, avec la TEI.

D'autre part l'édition électronique ne se substitue pas au document, elle n'en est pas une photographie, c'est pourquoi il est tout indiqué de fournir un lien vers l'image du manuscrit reproduit lorsqu'il est disponible.

Exemples d'éditions savantes voire critiques avec des perspectives différentes sur l'annotation critique :

- édition critique du texte du livre sur les poissons de l'Hortus sanitatis, avec traduction : <https://www.unicaen.fr/puc/sources/depiscibus/>
- édition critique de la Chronique de Geoffroi Malaterra : <https://www.unicaen.fr/puc/sources/malaterra/>

- édition sans appareil critique du texte d'encyclopédies médiévales latines, avec une couche d'annotation comprenant l'identification des sources : <http://sourcencyme.irht.cnrs.fr/>

(D) On essaiera d'éviter de tomber dans les pièges contre lesquels nous mettons en garde les auteurs de l'introduction des *Conseils pour l'édition des textes médiévaux*, soit le refus de prendre parti, assimilé ici à une paresse intellectuelle, mais ailleurs, comme chez Robinson, à une volonté un peu trop enthousiaste de rendre accessible tout le matériel existant ou du moins vu par l'éditeur, et, second piège, l'illusion qui consisterait à vouloir obtenir un rendu totalement fidèle à la source.

## (D) L'édition électronique, comment ?

### (D) Etapes du travail

#### **Phase d'acquisition du texte**

##### **Partir d'un document texte**

**méthode** : pré-balisage au moyen d'une feuille de style définissant des styles de caractères et des styles de paragraphe, ce qui permet de baliser deux niveaux de profondeur ou de granularité

##### **outils** :

- éditeur de texte (Word, LibreOffice) qui dispose de styles de caractère et de paragraphe
- feuilles de transformation XSLT qui vont permettre de passer d'un format à un autre, de transformer un prébalisage en balisage

##### **ressources** : Oxgarage

Inconvénients : ce prébalisage ne permet pas un encodage fin de l'apparat critique tel qu'on peut le développer avec d'autres moyens, comme les recommandations de la TEI dont je vais parler plus loin.

Cela peut cependant être un moyen commode de baliser facilement la structure du texte, de la transformer pour obtenir un fichier disposant déjà de grandes divisions ainsi que d'un encodage des éléments de premier niveau.

#### **(D) Utiliser un OCR ?**

Je ne vais pas vraiment développer cet aspect mais je renvoie ici vers les travaux et les réalisations des projets pilotés par Dominique Stutzmann ici à l'IRHT, que j'ai déjà mentionné mardi, ou l'environnement de traitement de manuscrits eScriptorium dirigé par Daniel Stökl Ben Ezra et Peter Stokes, en cours de développement, qui vise à fournir des outils pour la reconnaissance automatique de texte manuscrit. Reste ici la question du prébalisage ou du balisage du texte qui devra être effectué manuellement.

#### **(D) Encoder en TEI**

On peut aussi encoder directement en XML-TEI (je renvoie ici aux diapos consacrées à la TEI dans la présentation générale sur les humanités numériques).

Les recommandations de la TEI (*Text Encoding Initiative*) désignent un ensemble de règles destinées faciliter l'échange et le stockage d'informations électroniques, qui peuvent être principalement des textes, dans le domaine des SHS, mais aussi des images ou des sons. Il peut s'agir de ressources existantes ou de nouvelles ressources, que l'on crée (comme des transcriptions) en même temps qu'on les encode.

Le principe est de rendre explicites certaines caractéristiques des éléments d'un document afin de faciliter le traitement de ce document par des applications fonctionnant sur des systèmes différents. On appelle ce travail d'explicitation « balisage » ou « encodage »

Les recommandations de la TEI utilisent le langage XML (eXtensible Markup Language). XML est un langage de balises extensible, qui décrit la structure des documents et permet de séparer la le sens que l'on donne à l'information de l'apparence que l'on souhaite donner à cette information, ce qui représente une différence fondamentale par rapport à HTML par exemple, dont le rôle est principalement d'afficher les données. XML est aujourd'hui l'outil le plus répandu pour la transmission d'informations entre différentes applications.

XML permet, grâce à l'encodage du document, de définir à l'intérieur de ce document des **éléments**, des **attributs** et des règles qui régissent leur existence et leur agencement au sein du document. Les éléments fournissent des étiquettes à des séquences de texte, tandis que les attributs vont permettre de qualifier ou spécifier ces éléments.

XML permet de définir des espaces de noms, qui sont autant de vocabulaires normalisés, spécifiques à une utilisation particulière. Ces vocabulaires sont définis par des communautés d'utilisateurs. Ainsi, TEI correspond en quelque sorte à une application particulière d'XML.

C'est pourquoi n'importe quel éditeur capable de gérer des documents XML permet de créer ou de modifier des documents TEI, ce qui assure la pérennité des documents.

### L'apparat critique en XML-TEI

(D) Les recommandations de la TEI sont structurées autour de modules thématiques, dont voici la liste.

(D) Pour réaliser une édition critique, en plus des modules « de base », notamment pour la structure du texte, les entités nommées, la question des liens entre éléments, on va être amené à utiliser plusieurs autres modules, parmi lesquels en particulier :

(D) Le module 10. Manuscript description : permet de décrire et de cataloguer des manuscrits

Le module 11. Representation of primary sources : module contenant les éléments nécessaires pour réaliser des facsimile avec ou sans transcription

Enfin le module 12. Critical apparatus : module permettant d'encoder les variantes de plusieurs manuscrits

Les principes du module 12 sont comme suit. Ces manuscrits seront décrits dans l'en-tête du fichier, l'élément *teiHeader* sous forme d'une liste de témoins : liste de *witness* contenue dans un élément englobant *listWit*.

Chacun de ces témoins reçoit un attribut dont la valeur est un identifiant unique qui permet de cibler sans équivoque cet élément lorsque l'on en a besoin. Il s'agit de l'attribut *xml:id*, qui représente son sigle, auquel on fera référence dans le corps du texte, au sein de l'apparat.

Ces éléments *listWit* peuvent avoir du contenu en texte libre ou une description structurée de manuscrit avec l'élément *msDesc*, voire l'élément *bibl*.

(D) commenter

(D) L'apparat critique est encodé dans un élément *app*, qui contient des élément *rdg* pour les différentes variantes. Les leçons sélectionnées par l'éditeur peuvent être encodées non dans un *rdg*, mais dans un *lem*. *rdg* autant que *lem* sont dotés d'un attribut *wit*, qui rassemble des pointeurs vers les différents témoins encodés dans le *teiHeader*.

Les éléments *rdg* contenant les variantes peuvent recevoir des attributs *@type* et *@cause*. D'autre part, on peut marquer les variantes significatives au moyen d'autres attributs plus génériques, comme l'attribut *@n* qui est disponible sur tous les éléments.

(D) commenter – *div @type* - pas de *lem* ici

(D) 3 méthodes sont proposées par les Guidelines pour l'encodage de l'apparat critique :

• location-referenced method : l'encodage de l'apparat se fait hors du texte, et on renvoie de l'élément *app* vers le segment ou le bloc de texte concerné grâce à l'attribut *loc*. L'inconvénient est

que cette méthode n'est pas très précise.

- (D) exemple
- (D) double end-point attachment method : des ancrs (anchor) sont placées pour signaler le début et la fin du segment variant. Les leçons variantes sont encodées aussi à un autre endroit mais cette méthode est un peu plus précise puisqu'on peut délimiter le début et la fin du segment.
- (D) parallel segmentation method : on reste à l'intérieur du texte, et on encode directement un jeu de variantes. Cette méthode est très efficace mais peut se révéler difficile à mettre en œuvre dans les cas de variantes complexes.

On encode le type de méthode choisi dans l'en-tête du fichier, l'élément *teiHeader*.

### (D) Pourquoi encoder en TEI ?

- vocabulaire éprouvé depuis les années 90 et maintenu par une communauté très active et dynamique, dont il est assez improbable qu'elle faiblisse
- vocabulaire qui s'appuie sur un standard du World Wide Web Consortium : XML, donc même si TEI s'effondre et n'est plus maintenue, les fichiers seront toujours lisibles en XML
- vocabulaire spécifiquement développé et mis à jour au sein de modules spécifiques par des experts et des chercheurs de chaque domaine, qui autorise une finesse d'encodage extrême et adaptable à tous types de projets

(D) Les faiblesse, car il y en a quand même :

- la richesse du vocabulaire, qui est évidemment aussi son atout, implique qu'il y a toujours au moins 3 ou 4 manières différentes d'encoder un même phénomène textuel, ce qui génère des divergences d'encodage entre projets qui nuisent à l'idéal d'interopérabilité, dont on revient d'ailleurs un peu depuis quelques années, avec l'accumulation des expériences, c'est ce dont j'ai parlé.
- le fait de reposer sur un standard, XML, implique d'en respecter aussi les contraintes, et notamment l'interdiction des structures chevauchantes, ce qui pose des problèmes notamment lorsque l'on encode l'apparat critique en parallel segmentation et que l'on veut aussi annoter d'autres types d'éléments comme la terminologie, les noms etc. Les recommandations de la TEI fournissent des solutions pour remédier à ce problème, solutions dont la mise en œuvre est plus ou moins complexe et complexifie aussi l'exploitation du texte par la suite.

### Outils pour encoder en TEI

- (D) un éditeur XML quelconque :
  - Notepad ++, JEdit ; libres et gratuits mais n'offrent pas de possibilité de créer une interface de travail comme les deux autres
  - Oxygen : le plus utilisé car il fournit un support pour l'encodage en XML-TEI, il n'est pas excessivement cher à l'achat et les programmeurs sont très réactifs notamment sur le forum du site pour aider les utilisateurs dans la configuration avancée du programme
  - XMLMind XML Editor : propose le même genre de configuration même si un environnement sur XMLMind n'est pas transposable sur Oxygen directement et inversement, les technologies sont un peu différentes ; à noter que il est proposé par une boîte française ; l'autre caractéristique est qu'il est davantage centré sur les questions de validation :
    - un fichier valide est un fichier qui respecte un schéma : le schéma TEI-All qui est le schéma le plus large disponible avec tous les éléments TEI ou un schéma réduit par exemple à l'aide de l'outil Roma
    - à ne pas confondre avec un fichier bien formé, qui respecte les règles d'XML
    - cette question de la validation est importante, j'ai parlé de modélisation des données

dans la séance plénière : si ce travail n'est pas fait en amont ou qu'il n'est pas respecté, c'est-à-dire que l'encodage de votre fichier ne respecte pas la modélisation qui a été faite au début, vous allez avoir des problèmes pour exploiter votre fichier dans la suite du processus, c'est pourquoi certaines équipes préfèrent travailler avec XMLMind qui est plus strict de ce point de vue parce que cela permet de contraindre la saisie et d'éviter de sortir du schéma

## **(D) Publier son édition**

### **Pour obtenir un apparat critique « papier »**

#### Éditeur de texte

avec notes de bas de page

#### LaTeX

LaTeX est langage de composition qui est utilisé pour la mise en page. Deux modules entre autres ont été développés pour l'utilisation par les philologues :

- reledmac : pour la composition de l'apparat critique accompagné de plusieurs étages d'apparat, avec notamment le calcul du numéro de ligne à reporter dans l'apparat
- reledpar : pour la publication d'éditions synoptiques, par exemple le texte et sa traduction

(D) J'ai utilisé ces deux modules dans mon édition du livre 3 du CP dans ma thèse en mettant en regard les 4 versions, chacune avec son apparat, plus un apparat d'identification des sources.

#### CTE

Je cite le Classical Text Editor avec lequel plusieurs éditions critiques ont été réalisées et qui permet d'avoir également plusieurs étages de notes, mais dont le prix reste assez prohibitif.

#### (D) Ponts entre TEI et LaTeX

L'intérêt est ici évidemment de développer des ponts entre XML-TEI et LaTeX, pour espérer générer à la fois l'édition critique papier et l'édition électronique à partir du même fichier XML-TEI. On va utiliser pour cela des feuilles de transformation XSLT pour passer d'un format à un autre.

(D) Il existe des outils en ligne pour réaliser ce genre de transformation, notamment la TEI Critical Apparatus Toolbox de M. Burghart, dont la feuille de transformation est un peu générique mais peut être adaptée. Cet outil permet de transformer une édition critique encodée en TEI Parallel segmentation en édition papier (PDF) : <http://teicat.huma-num.fr/print.php>

### **(D) Pour réaliser une édition électronique**

#### À partir d'un fichier XML-TEI : vers XHTML

Là encore on va utiliser des feuilles de transformation XSLT pour générer des pages web à partir d'un encodage en XML-TEI.

Encore une fois des outils génériques existent comme le moteur d'affichage Max développé par le pôle « Document numérique » de la MRSH de Caen : [https://www.unicaen.fr/recherche/mrsh/document\\_numerique/outils/max](https://www.unicaen.fr/recherche/mrsh/document_numerique/outils/max). Mais ce moteur doit vraiment être adapté à chaque projet avec l'aide de collègues informaticiens.

#### Ressources

**les guidelines de la TEI**, la version web, qui sont une vraie bible pour qui veut s'engager dans cette voie

**Méthodologies** : peu pour l'encodage de l'apparat critique, il en existe pour la description des

reliures à la BNF, pour l'encodage d'inventaires anciens (publié au sein des outils mis à disposition par l'Equipex Bibliissima)

**Liste de diffusion TEI-L**, avec une communauté très active

**Aides à l'encodage** : environnements de balisage sous :

- XMLMindXML Editor
- ou sous oXygen, qui permettent plusieurs choses :
  - afficher le document sous des vues différentes, au moyen de feuilles de style CSS, dont chacune va décrire la façon que le logiciel va avoir d'afficher chaque élément : on peut donc par exemple avoir :
    - une vue de travail dans laquelle on va encoder les variantes
    - une vue de lecture de chaque manuscrit reconstitué, si l'on travaille avec la méthode Parallel segmentation
    - une vue du texte édité au moyen des éléments lem avec un simili appareil rudimentaire
  - automatiser un grand nombre de tâches, notamment :
    - la référence à des métadonnées décrites dans l'en-tête, comme la référence à des témoins au sein des éléments rdg pour les variantes, la référence à des mains également décrites dans l'en-tête
    - le renseignement de valeurs d'attributs avec des listes fermées qui peuvent apparaître dans des listes déroulantes
    - (ex. avec XXE) la semi-automatisation de la référence à des éléments externes comme des listes d'autorités ou des éléments contenus dans un autre fichier

Autres outils et ressources :

- boîte à outils pour l'historien des textes : <https://baobab.bibliissima.fr/>
- répertoire de références et ressources pour le médiéviste : <http://www.menestrel.fr/>
- outil de transcription d'images : <https://transkribus.eu/lite/fr>
- outil de collation et de création d'apparat, pour tout type de langue (y compris RTL) : <https://cental.uclouvain.be/chrysocollate/>
- TEI Critical apparatus toolbox avec divers outils pour l'édition en TEI Parallel segmentation : <http://teicat.huma-num.fr/>
- application pour transformations avec feuilles XSLT génériques : <https://oxgarage.tei-c.org/>
- ressources pour l'édition de textes arabes ou multilingues : <https://www.robtalessi.net/>

## (D) L'édition électronique, avec qui ?

Les outils en cours de développement ou déjà disponibles visent à permettre au chercheur d'encoder lui-même son édition et donc de maîtriser de bout en bout la chaîne de production de l'édition critique. Cependant la réussite du processus dépend aussi d'une collaboration avec des personnes capables d'orienter le chercheur sur les choix techniques et notamment sur le choix du vocabulaire XML-TEI, qu'on appelle « schéma », en fonction des objectifs du projet éditorial, mais aussi capables de réaliser la transformation vers d'autres formats et notamment le pdf pour le format papier et l'XHTML pour le format électronique. Cela reste donc un travail mené de bout en bout par le chercheur, tout en fonctionnant en collaboration avec des informaticiens.

Important :

- se faire accompagner
- ne pas se lancer seul dans l'encodage !
- suivre des formations
- la publication électronique demandera de vraies compétences informatiques