

EXEMPLIER**Table des matières**

Exercice 1 : Structurer un texte et créer des index.....	1
Consignes de l'exercice.....	1
Encoder la structure du document.....	1
Création d'un index des lieux et des noms de personne.....	2
2.4 The Profile Description.....	3
13.3.4 Places.....	4
13.3.2 Person.....	4
Exercice 2 : Encoder des variantes textuelles.....	5
Consignes de l'exercice.....	5
12-Critical Apparatus.....	6
12.1.1 The Apparatus Entry.....	6
12.1.4.3 The Witness List.....	7
12.2 Linking the Apparatus to the Text.....	7
12.2.3 The Parallel Segmentation Method.....	7

Exercice 1 : Structurer un texte et créer des index***Consignes de l'exercice***

Nous allons encoder les deux premiers paragraphes de *Notre Dame de Paris*, II, chapitre 1, «De Charybde en Scylla» (début de chapitre → «La cohue admirait»). Nous transcrivons le plan de l'œuvre sur Oxygène. Puis nous repèrerons dans l'extrait les noms de lieux afin d'en réaliser un index.

Pour que votre réalisation soit un outil utile, il faudrait bien entendu réaliser cela pour l'ensemble du texte, ce que vous pourrez faire en autonomie sur votre corpus.

Objectifs :

- comprendre l'arborescence du XML
- utiliser les balises comme éléments d'analyse
- utiliser les balises comme exploitation de données, en générant un index.

Un exemplaire est à votre disposition, mais la structure du livre est aussi disponible sur wikisource :

[https://fr.wikisource.org/wiki/Notre-Dame de Paris/Livre deuxieme#II. La place de Gr.C3.A8ve](https://fr.wikisource.org/wiki/Notre-Dame_de_Paris/Livre_deuxieme#II._La_place_de_Gr.C3.A8ve)

Encoder la structure du document.

- a) Ouvrir un nouveau document XML sur Oxygène : Nouveau document → Dossier TEI P5 → All. Le document qui apparaît possède déjà une structure valide qu'il s'agira de remplir et respecter. Les premières lignes montrent qu'il a un schéma qui respecte les exigences de la TEI P5.
- b) Remplir le Teiheader avec les informations sur le texte.
- c) Pour la partie texte, on ne remplira que les informations qui nous permettent d'encoder cet extrait. Donc on peut oublier la préface, et n'écrire que le titre du livre 1, sans les chapitres à l'intérieur.

Quand le texte est déjà disponible au format numérique, un simple copier-coller suffit. Il faut néanmoins nettoyer le texte : enlever les doubles espaces, les tableaux, la couleur, tout ce qui a pu rester de la typographie du site pour avoir un texte le plus brut possible.

La nuit arrive de bonne heure en janvier. Les rues étaient déjà sombres quand Gringoire sortit du Palais. Cette nuit tombée lui plut ; il lui tardait d'aborder quelque ruelle obscure et déserte pour y méditer à son aise et pour que le philosophe posât le premier appareil sur la blessure du poète. La philosophie était du reste son seul refuge, car il ne savait où loger. Après l'éclatant avortement de son coup d'essai théâtral, il n'osait rentrer dans le logis qu'il occupait, rue Grenier-sur-l'Eau, vis-à-vis le Port-au-Foin, ayant compté sur ce que monsieur le prévôt devait lui donner de son épithalame pour payer à maître Guillaume Doulx-Sire, fermier de la coutume du pied-fourché de Paris, les six mois de loyer qu'il lui devait, c'est-à-dire douze sols parisis ; douze fois la valeur de ce qu'il possédait au monde, y compris son haut-de-chausses, sa chemise et son bicoquet. Après avoir un moment réfléchi, provisoirement abrité sous le petit guichet de la prison du trésorier de la Sainte-Chapelle, au gîte qu'il élirait pour la nuit, ayant tous les pavés de Paris à son choix, il se souvint d'avoir avisé, la semaine précédente, rue de la Savaterie, à la porte d'un conseiller au parlement, un marche-pied à monter sur mule, et de s'être dit que cette pierre serait, dans l'occasion, un fort excellent oreiller pour un mendiant ou pour un poète. Il remercia la providence de lui avoir envoyé cette bonne idée ; mais, comme il se préparait à traverser la place du Palais pour gagner le tortueux labyrinthe de la Cité, où serpentent toutes ces vieilles sœurs, les rues de la Barillerie, de la Vieille-Draperie, de la Savaterie, de la Juiverie, etc., encore debout aujourd'hui avec leurs maisons à neuf étages, il vit la procession du pape des fous qui sortait aussi du Palais et se ruait au travers de la cour, avec grands cris, grande clarté de torches et sa musique, à lui Gringoire. Cette vue raviva les écorchures de son amour-propre ; il s'enfuit. Dans l'amertume de sa mésaventure dramatique, tout ce qui lui rappelait la fête du jour l'aigrissait et faisait saigner sa plaie.

Il voulut prendre le pont Saint-Michel ; des enfants y couraient çà et là avec des lances à feu et des fusées.

— Peste soit des chandelles d'artifice ! dit Gringoire, et il se rabattit sur le pont au Change. On avait attaché aux maisons de la tête du pont trois drapeaux représentant le roi, le dauphin et Marguerite de Flandre, et six petits drapelets où étaient pourtraicts le duc d'Autriche, le cardinal de Bourbon, et M. de Beaujeu, et madame Jeanne de France, et M. le bâtard de Bourbon, et je ne sais qui encore ; le tout éclairé de torches. La cohue admirait.

- d) Vous pouvez basculer en mode auteur pour voir ce que donne votre encodage. Revenez au mode texte. Pour obtenir une **indentation** du texte qui le rendra plus lisible, il faut appuyer sur le bouton « formater et indenter ». Cela ne change rien à l'encodage contrairement au fait d'ajouter des sauts de ligne.

Création d'un index des lieux et des noms de personne.

- a) Lire le texte et repérer les noms de lieux. Les entourer de la balise <placeName>.
- b) Créer dans le Teiheader une liste des lieux. Pour cela, juste après la balise fermante </filedesc> on va ouvrir des balises au même niveau appelées <profiledesc></profiledesc>. Pour voir ce qu'on insère dans ces balises, rdv sur les guidelines à votre disposition. Pour voir un exemple d'emploi de ces balises, vous pouvez essayer de les repérer sur le document xml.base.
- c) Le premier lieu du texte est le « Palais », une forme abrégée de Palais de Justice. Le premier la balise <place> portera l'attribut xml:id = « Palais_de_Justice » (vous pouvez tester l'emploi des espaces et des apostrophes, la machine n'aime pas du tout ça sur les attributs !). Quel que soient les mots employés pour désigner le Palais de Justice dans le texte, il s'agira de les renvoyer à ce xml:id.

Dans la balise <placeName>, vous pouvez écrire le vrai nom, sans _ puis qu'il s'agit de texte.

Dans la balise <note> vous pouvez ajouter des informations sur le lieu.

Certaines balises peuvent être ajoutées pour préciser le positionnement géographique (latitude, longitude).

- d) A présent, il faut retourner au texte et ajouter au premier <placeName> concernant le Palais de Justice l'attribut : ref= « #Palais_de_Justice ». Normalement, Oxygen vous le propose dans une petite

fenêtre pop-up.

Le # est utilisé pour créer un pont entre la référence et la carte d'identité du lieu créé précédemment dans le header. On appelle cela un **pointeur**.

Chaque fois que vous trouverez ce lieu, vous pourrez réécrire cet attribut `ref= « #Palais_de_Justice. »`

- e) Reproduisez cette méthode pour les autres lieux du texte. (Si le temps vous manque, vous pourrez à titre exceptionnel en oublier quelques-uns).
- f) Pour aller plus loin : faire un index des noms de personnages. En trouvant comment insérer la balise `<person>` dans le header d'après les TEI Guidelines.

Vous pourrez choisir d'indexer tous les noms propres ou bien seulement ceux des personnages. Tout dépend de l'utilisation que vous comptez faire de votre index (faciliter la lecture d'extraits ou proposer un outil linguistique par exemple).

2.4 The Profile Description

The `profileDesc` element is the third major subdivision of the TEI header. It is an optional element, the purpose of which is to enable information characterizing various descriptive aspects of a text or a corpus to be recorded within a single unified framework.

profileDesc (description du profil) fournit une description détaillée des aspects non bibliographiques du texte, notamment les langues utilisées et leurs variantes, les circonstances de sa production, les collaborateurs et leur statut.

The `profileDesc` element contains elements taken from the `model.profileDescPart` class. The default members of this class are the following :

- `abstract` contains a summary or formal abstract prefixed to an existing source document by the encoder.
- `creation` (création) contient des informations concernant la création d'un texte.
- `langUsage` (langue utilisée) décrit les langues, variétés de langues, registres, dialectes, etc. présents à l'intérieur d'un texte.
- `textClass` (classification du texte) regroupe des informations décrivant la nature ou le sujet d'un texte selon des termes issus d'un système de classification standardisé, d'un thésaurus, etc.
- `correspDesc` (correspondence description) contains a description of the actions related to one act of correspondence.
- `calendarDesc` (description des calendriers) contient la description des différents calendriers utilisés dans des dates écrits dans un manuscrit.

These elements are further described in the remainder of this section.

When the corpus module described in chapter 15 Language Corpora is included in a schema, three further elements become available within the `profileDesc` element:

- `textDesc` (description de texte) fournit la description d'un texte sous l'angle du contexte situationnel

- **particDesc** (description des participants) décrit les locuteurs, voix ou autres participants identifiables d'une interaction linguistique.
- **settingDesc** (description du contexte) décrit le(s) contexte(s) dans lesquels se situe une interaction linguistique, soit sous la forme d'une description en prose, soit sous celle d'une série d'éléments décrivant le contexte.¹

13.3.4 Places

In [13.2.3 Place Names](#) we discuss various ways of naming places such as towns, countries, etc. In much the same way as these Guidelines distinguish between the encoding of names for people and the encoding of other data about people, so they also distinguish between the encoding of names for places and the encoding of other data about places. In this section we present elements which may be used to record in a structured way data about places of any kind which might be named or referenced within a text. Such data may be useful as a way of normalizing or standardizing references to particular places, as the raw material for a gazetteer or similar reference document associated with a particular text or set of texts, or in conjunction with any form of geographical information system.²

The following elements are provided for this purpose:

- [listPlace](#) (liste de lieux) contient une liste de lieux, qui peut être suivie d'une liste de relations définies entre les lieux (autres que la relation d'inclusion).
- [place](#) (lieu) contient des informations sur un lieu géographique.

```
<settingDesc>
<listPlace type="offshoreIslands">
  <place>
    <placeName>La roche qui pleure</placeName>
  </place>
  <place>
    <placeName>Ile aux cerfs</placeName>
  </place>
</listPlace>
</settingDesc>
```

Within the [place](#) element, the following elements may be used to provide more information about specific aspects of the place in a structured form:

- [placeName](#) (nom de lieu) contient un nom de lieu absolu ou relatif.
- [location](#) (localisation) définit l'emplacement d'un lieu par des coordonnées géographiques, en termes d'entités nommées dites géopolitiques, ou par une adresse.

13.3.2 Person

Information about a person, as distinct from references to a person, for example by name, is grouped together within a person element. Information about a group of people regarded as a single entity (for example 'the audience' of a performance) may be encoded using the personGrp element. Note however that information about a group of people with a distinct identity (for example a named theatrical troupe) should be recorded using the org element described in section 13.3.3 Organizational Data below.

¹ Explications extraites des TEI Guidelines : <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/HD.html#HD4>

² Explications extraites des TEI Guidelines : <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html#NDGEOG>

These elements may appear only within a listPerson element, which groups such descriptions together, and optionally also describes relationships amongst the people listed.

- listPerson (liste de personnes) contient une liste d'éléments, chacun d'entre eux apportant des informations sur une personne précise ou sur un groupe de personnes, par exemple les participants à une interaction linguistique, ou les personnes citées dans une source historique.
- listRelation donne des informations sur les relations qui existent entre des personnes, des lieux, ou des organisations, soit de manière informelle en prose, soit de manière formelle.

One or more listPerson elements may be supplied within the particDesc (participant description) element in the profileDesc element of a TEI header (see 2.4 The Profile Description). Like other forms of list, however, the listPerson can also appear within the body of a text when the module defined by this chapter is included in a schema.³

```
<profileDesc>
  <particDesc>
    <listPerson type="historical">
      <person xml:id="ART1">
        <persName>Arthur</persName>
      </person>
      <person xml:id="BERT1">
        <persName>Bertrand</persName>
      </person>
    <!-- ... -->
    </listPerson>
    <listPerson type="mythological">
      <person xml:id="ART2">
        <persName>Arthur</persName>
      </person>
      <person xml:id="BERT2">
        <persName>Bertrand</persName>
      </person>
    <!-- ... -->
    </listPerson>
  </particDesc>
</profileDesc>
```

Exercice 2 : Encoder des variantes textuelles

Consignes de l'exercice

Création d'une édition numérique — niveau moyen —

Vous avez trois versions d'un texte, vous devez encoder le texte proposé en l'organisant en différentes sections avec un appareil critique selon les principes de la *parallel-segmentation* qui permettra de montrer à terme les différentes versions du texte.

Une fois que vous aurez réalisé votre fichier XML avec un teiHeader renseigné, vous essayerez de le transformer en vous aidant de la toolbox, outil développé par Marjorie Burgarth au CIHAM.

- lien vers la toolbox : <http://teicat.huma-num.fr>

Faire l'exercice pas à pas :

- 1- Créer un nouveau fichier XML-TEI dans Oxygen – choisir la version TEI P5 all ;
- 2- Aller dans le dossier drive partagé / ouvrir le fichier *troisVersionsConte* / copier dans le <body> du fichier XML la version qui servira de base de votre encodage ;

³ Explications extraites des TEI Guidelines :
<http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html#NDPERSE>

- 3- Renseigner le `teiHeader`, sans oublier de renseigner les métadonnées sur vos témoins pour lesquels vous utiliserez la balise `<listWit>` qui est la plus synthétique ;
- 4- Structurer votre texte en paragraphe à l'aide de la balise `<p>` et de la balise `<said>` pour les dialogues ;
- 5- Proposer un encodage des différentes versions du conte en suivant les prescriptions d'encodage d'un apparat selon la méthode : *parallel-segmentation*. – N'oubliez pas de déclarer votre méthode dans le `TeiHeader` –
- 6- Quand vous aurez terminé votre encodage, sauvegardez votre document ;
- 7- Suivre le lien vers la toolbox ;
- 8- Cliquer sur l'onglet : *Check your encoding*, charger votre fichier sur la page puis soumettez-le. Observez les résultats obtenus ;
- 9- Cliquer sur l'onglet : *Display parallel versions*, charger votre fichier sur la page, puis soumettez-le. Observer les résultats obtenus.

12-Critical Apparatus

Scholarly editions of texts, especially texts of great antiquity or importance, often record some or all of the known variations among different witnesses to the text. Witnesses to a text may include authorial or other manuscripts, printed editions of the work, early translations, or quotations of a work in other texts. Information concerning variant readings of a text may be accumulated in highly structured form in a critical apparatus of variants. This chapter defines a module for use in encoding such an apparatus of variants, which may be used in conjunction with any of the modules defined in these Guidelines. It also defines an element class which provides extra attributes for some elements of the core tag set when this module is selected. In printed critical editions, the apparatus takes the form of highly-compressed notes at the bottom of each page. TEI's critical apparatus module allows variation to be encoded so that such notes may be generated, but it also models the variation so that, for example, interactive editions in which readers can choose which witness readings to display are possible.

Information about variant readings (whether or not represented by a critical apparatus in the source text) may be recorded in a series of apparatus entries, each entry documenting one variation, or set of readings, in the text. Elements for the apparatus entry and readings, and for the documentation of the witnesses whose readings are included in the apparatus, are described in section [12.1 The Apparatus Entry, Readings, and Witnesses](#). Special tags for fragmentary witnesses are described in section [12.1.5 Fragmentary Witnesses](#). The available methods for embedding the apparatus in the rest of the text, or for linking an external apparatus to the base text, are described in section [12.2 Linking the Apparatus to the Text](#). Finally, several extra attributes for some tags of the core tag set, made available when the additional tag set for text criticism is selected, are documented in section [11.3.1.1 Core Elements for Transcriptional Work](#).⁴

12.1.1 The Apparatus Entry

Individual textual variations are encoded using the `app` element, which groups together all the readings constituting the variation. The identification of discrete textual variations or apparatus entries is not a purely mechanical process; different editors may group readings differently. No rules are given here as to how to group readings into apparatus entries; the tags given here may be used to group readings in whatever way the editor finds most perspicuous or useful.

The individual apparatus entry is encoded with the `app` element:

4 Explications extraites des TEI Guidelines : <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/TC.html>

- [app](#) (entrée d'apparat critique) contient une entrée dans un appareil critique, constituée d'un lemme facultatif et d'au moins une leçon.

```
<app>
<rdg wit="#El">Experience though noon Auctoritee</rdg>
<rdg wit="#La">Experiment thouh noon Auctoritee</rdg>
<rdg wit="#Ra2">Eryment though none auctorte</rdg>
</app>
```

12.1.4.3 The Witness List

A list of all identified witnesses should normally be supplied in the front matter of the edition, or in the [sourceDesc](#) element of its header. This may be given either as a simple bibliographic list, using the [listBibl](#) element described in [3.11 Bibliographic Citations and References](#), or as a [listWit](#) element, which contains a series of [witness](#) elements. Each [witness](#) element may contain a brief characterization of the witness, given as one or more prose paragraphs. If more detailed information about a manuscript witness is available, it should be represented using the [msDesc](#) element provided by the msdescription module; a [msDesc](#) may appear within a [listBibl](#).

Whether information about a particular witness is supplied by means of a [bibl](#), [msDesc](#), or [witness](#) element, a unique sigil for this source should always be supplied, using the global `xml:id` attribute. This identifier can then be used elsewhere to refer to this particular witness.

- [listWit](#) (liste de témoins) donne une liste de définitions pour tous les témoignages cités dans un appareil critique, pouvant être groupées de façon hiérarchique.
- [witness](#) (témoin) contient soit la description d'un seul témoin auquel il est fait référence à l'intérieur de l'apparat critique, soit une liste de témoins, à laquelle on doit faire référence par une seule abréviation
- [msDesc](#) (description d'un manuscrit) contient la description d'un manuscrit individuel
- [bibl](#) (référence bibliographique) contient une référence bibliographique faiblement structurée dans laquelle les sous-composants peuvent ou non être explicitement balisés.
- [listBibl](#) (liste de références bibliographiques) contient une liste de références bibliographiques de toute nature.

```
<listWit>
<witness xml:id="El">Ellesmere, Huntingdon Library 26.C.9</witness>
<witness xml:id="Hg">Hengwrt, National Library of Wales,
  Aberystwyth, Peniarth 392D</witness>
<witness xml:id="Ra2">Bodleian Library Rawlinson Poetic 149</witness>
</listWit>
```

12.2 Linking the Apparatus to the Text

Three different methods may be used to link a critical apparatus to the text:

- 1.the location-referenced method,
- 2.the double-end-point-attached method, and
- 3.the parallel segmentation method.

Both the location-referenced and the double end-point methods may be used with either in-line or external apparatus, the former dispersed within the base text, the latter held in some separate location, within or outside the document with the base text. The parallel segmentation method does not use the concept of a base text and may only be used for in-line apparatus.

Any document containing [app](#) elements requires a [variantEncoding](#) declaration in the [encodingDesc](#) element of its TEI header, thus:

- [variantEncoding](#) (Méthode d'encodage des variantes) sert à déclarer la méthode utilisée pour encoder les variantes critiques du texte.

12.2.3 *The Parallel Segmentation Method*

This method differs from the double end-point attachment method in that all variants at any point of the text are expressed as variants on one another. In this method, no two variations can overlap, although they may nest. The texts compared are divided into matching segments all synchronized with one another. This permits direct comparison of any span of text in any witness with that in any other witness. It is also very easy with this method for an application to extract the full text of any one witness from the apparatus.

This method will (by definition) always be satisfactory when there are just two texts for comparison (assuming they are in the same language and script). It will however be less convenient for very complex traditions, where establishing a base text with variations from it is not a satisfactory goal for the edition, or in some cases where every detail of variation needs to be modeled.

In the parallel segmentation method, each segment of text on which there is variation is marked by an [app](#) element. If there is a preferred (or base) reading it is tagged with [lem](#); each reading is given in a [rdg](#) element:

```
<variantEncoding method="parallel-segmentation"
location="internal"/>
<!-- ... -->
<| n="1">
  <app>
    <rdg wit="#El #Hg">Experience</lem>
    <rdg wit="#La">Experiment</rdg>
    <rdg wit="#Ra2">Eryment</rdg>
  </app>
  though noon Auctoritee
</|>
<|>Were in this world ...</|>
```