

Sarrasine de Balzac, balisage personnel vs balisage respectueux de la TEI:P5

Présentation de la DTD proposée et des éléments TEI correspondants :

DTD personnelle	Éléments de la TEI:P5
<!ELEMENT document (titre, Npersonne+, dédicace?, texte, colophon?)> <!ATTLIST document édition CDATA #REQUIRED>	<text> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-text.html : « <text> (texte) contient un seul texte quelconque, simple ou composite, par exemple un poème ou une pièce de théâtre, un recueil d'essais, un roman, un dictionnaire ou un échantillon de corpus. [4 Default Text Structure 15.1 Varieties of Composite Text] »
<!-- les 3 premiers fils de <document> (titre, Npersonne, dédicace) constituent un paratexte antéposé -->	<front> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-front.html : « <front> (texte préliminaire) contient tout ce qui est au début du document, avant le corps du texte : page de titre, dédicaces, préfaces, etc. [4.6 Title Pages 4 Default Text Structure] »
<!ELEMENT titre (#PCDATA)>	<docTitle> contenant <titlePart> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-docTitle.html : « <docTitle> (titre du document) contient le titre d'un document, incluant la totalité de ses composants tels qu'ils sont donnés sur la page de titre. [4.6 Title Pages] » cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-titlePart.html : « <titlePart> contient une section ou division du titre d'un ouvrage telle qu'elle est indiquée sur la page de titre. [4.6 Title Pages] »
<!ELEMENT Npersonne (#PCDATA choix)*> <!ATTLIST Npersonne type (auteur dédicataire autre) #IMPLIED> <!-- l'élément <Npersonne> fils de <document> et placé entre les éléments <titre> et <dédicace> porte l'@type="auteur" -->	<docAuthor> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-docAuthor.html : « <docAuthor> (auteur du document) contient le nom de l'auteur du document tel qu'il est donné sur la page de titre (ce nom est le plus souvent contenu dans une mention de responsabilité) . [4.6 Title Pages] »
<!ELEMENT choix (Houssiaux?, Furne?) >	<choice> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-choice.html : « <choice> (choix) regroupe un certain nombre de balisages alternatifs possibles pour un même endroit dans un texte. [3.4 Simple Editorial Changes] »
<!ELEMENT Houssiaux (#PCDATA) >	<orig> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-orig.html : « <orig> (forme originale) contient une partie notée comme étant fidèle à l'original et non pas normalisée ou corrigée. [3.4.2 Regularization and Normalization 12 Critical Apparatus] »
<!ELEMENT Furne (#PCDATA) >	<reg> cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-reg.html : « <reg> (régularisation) contient une partie qui a été régularisée ou normalisée de façon quelconque [3.4.2 Regularization and Normalization 12 Critical

DTD personnelle	Éléments de la TEI:P5
<!ELEMENT dédicace (#PCDATA Npersonne choix)*>	<p>Apparatus »</p> <p><div> @type="dedication" contenant <salute></p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-div.html : « <div> (division du texte) contient une subdivision dans le texte préliminaire, dans le corps d'un texte ou dans le texte postliminaire. [4.1 Divisions of the Body] »</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-salute.html : « <salute> (formule de politesse) contient une dédicace ou une formule de salut qui précède un avant-propos ou autre division du texte; ou bien encore la formule de politesse qui conclut une lettre, une préface, etc. [4.2.2 Openers and Closers] »</p>
<!ELEMENT texte ((p dialogue)+)>	<p><body></p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-body.html : « <body> (corps du texte) contient la totalité du corps d'un seul texte simple, à l'exclusion de toute partie pré- ou post-liminaire. [4 Default Text Structure] »</p>
<!ELEMENT p (#PCDATA Npersonne Npersonnage Ntemps Nlieu NobjCulturel choix)*>	<p><p></p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-p.html : « <p> (paragraphe) marque les paragraphes dans un texte en prose. [3.1 Paragraphs 7.2.5 Speech Contents] »</p>
<p><!-- <!ELEMENT Npersonne (#PCDATA choix)* <!ATTLIST Npersonne type (auteur dédicataire autre) #IMPLIED --></p> <p><!-- Les éléments <Npersonne> qui ne sont pas fils de <document> et placés entre les éléments <titre> et <dédicace> porte un @type dont la valeur n'est pas "auteur" --></p>	<p><persName> @type="person"</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-persName.html : « <persName> (nom de personne) contient un nom propre ou une expression nominale se référant à une personne, pouvant inclure tout ou partie de ses prénoms, noms de famille, titres honorifiques, noms ajoutés, etc. [13.2.1 Personal Names] »</p>
<!ELEMENT Npersonnage (#PCDATA choix)* >	<p><persName> @type="character"</p>
<!ELEMENT Ntemps (#PCDATA choix)* >	<p><time> ou <date> (selon pertinence)</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-time.html : « <time> (temps) contient une expression qui précise un moment de la journée sous n'importe quelle forme. [3.5.4 Dates and Times] »</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-date.html : « <date> (date) contient une date exprimée dans n'importe quel format. [3.5.4 Dates and Times 2.2.4 Publication, Distribution, Licensing, etc. 2.6 The Revision Description 3.11.2.4 Imprint, Size of a Document, and Reprint Information 15.2.3 The Setting Description 13.3.7 Dates and Times] »</p>
<!ELEMENT Nlieu (#PCDATA choix)* >	<p><placeName> ou <geogName> (selon pertinence)</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-placeName.html : « <placeName> (nom de lieu) contient un nom de lieu absolu ou relatif. [13.2.3 Place Names] »</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-geogName.html : « <geogName> (nom de lieu géographique) un nom associé à une caractéristique géographique comme Windrush Valley ou le Mont Sinai.</p>

DTD personnelle	Éléments de la TEI:P5
	[13.2.3 Place Names] »
<!ELEMENT NobjCulturel (#PCDATA choix)* >	<p><title> ou <name> (selon pertinence)</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-title.html : « <title> (titre) contient le titre complet d'une oeuvre quelconque [3.11.2.2 Titles, Authors, and Editors 2.2.1 The Title Statement 2.2.5 The Series Statement] »</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-name.html : « <name> (nom, nom propre) contient un nom propre ou un syntagme nominal. [3.5.1 Referring Strings] »</p>
<!ELEMENT dialogue (TourDeParole+) >	<div>
<!ELEMENT TourDeParole (#PCDATA Npersonne Npersonnage Ntemps Nlieu NobjCulturel choix)* >	<p><said> ou <u> (au choix)</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-said.html : « <said> (Parole ou pensée.) indique que les passages sont pensés ou qu'ils sont prononcés à haute voix, que cela soit indiqué explicitement ou non dans la source , que ces passages soient directement ou indirectement rapportés par des personnages réels ou fictifs. [3.3.3 Quotation] »</p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-u.html : « <u> (énonciation) partie de discours généralement précédée et suivie d'un silence ou d'un changement de locuteur. [8.3.1 Utterances] »</p>
<!ELEMENT colophon (#PCDATA Ntemps Nlieu choix)* > <!-- comme les 3 premiers fils de <document> (titre, Npersonne, dédicace) constituent un paratexte antéposé, le dernier appartient à un paratexte postposé -->	<p><back> contenant <colophon></p> <p>cf. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-colophon.html : « <colophon> (colophon) contient le colophon d'une section d'un manuscrit, c'est-à-dire la transcription des informations relatives à la date, au lieu, à l'organisme commanditaire ou aux raisons de la production du manuscrit. [10.6.1 The msItem and msItemStruct Elements] »</p>

1. Balisage personnel de *Sarrasine* [20 points]

[Travail en groupe de 1 à 3 étudiants pour le 18/12/2020]

Faire le balisage minutieux du texte intégral de *Sarrasine* contenant

- l'annotation des différences entre les éditions Houssiaux *vs* Furne ;
- le repérage des éléments structurels (ceux vus en cours au moins) ;
- le balisage des noms d'entités (personnes, personnages, œuvres d'art diverses – dits objets culturels –, lieux et indications temporelles – peut-être en les révisant en fonction des distinctions faites dans la *TEI*) ;
- et, si vous le souhaitez, des enrichissements personnels relatifs, par exemple, à la distinction des récits enchâssant et enchâssé, à des champs lexicaux, aux emplois de mot empruntés à d'autres langues, aux noms d'autres entités (divinités...), etc.

Vous élaborerez donc une DTD basée sur les déclarations du tableau ci-dessus mais adaptée à vos besoins.

Pour réaliser ce balisage, travailler dans un traitement de texte avec des styles de paragraphes et de caractères (à augmenter si besoin par rapport à ce que nous avons fait en cours) puis convertir le document ODT obtenu en un fichier XML qu'il faudra réviser pour qu'il se conforme à la DTD.

Si le travail de balisage est partagé entre plusieurs étudiants, il faudra que les balisages soient cohérents et que la partie balisée par chacun soit clairement bornée au moyen de commentaires écrits sur les patrons suivants : « <!-- début du balisage par NOM_1 --> », puis « <!-- fin du balisage par NOM_1 et début du balisage par NOM_2 --> » et « <!-- fin du balisage par NOM_2 --> ».

Méthodologie pour un traitement ODT vers XML :

↓ stylage complet avec préservation des différences entre les éditions Houssiaux et Furne

*Sarrasine_***nom(s)***.odt*

↓ copie du fichier et changement de l'extension de la copie

*Sarrasine_***nom(s)***.zip*

↓ dézippage dans un dossier de même nom que le ZIP

*Sarrasine_***nom(s)***> content.xml*

↓ ouverture du fichier dans <oXygen/>

↓ passage du mode éditeur au mode débogueur XSLT (icône en haut à droite)

↓ transformation avec *ODTversXML=Audrain+Gasiglia_ODT2017-02.xslt*

↓ enregistrement du produit de la transformation (qui est affiché dans la zone droite du débogueur)

*Sarrasine_de_ODT_***nom(s)***.xml*

↓ transformations avec *XML_vers_XML=suppr_Ti.xsl* (éventuellement enrichie d'autres noms d'éléments "parasites" à supprimer)

*Sarrasine_***nom(s)***.xml*

↓ suppression des préfixes « UDL » des noms de styles ODT devenus noms d'éléments (sélectionné UDL, utilisé les touches Ctrl F (ou Cmd F) pour ouvrir la fenêtre de chercher/remplacer, et demander d'activer les options de recherche XML en sélectionnant que nos remplacements touchent des noms d'éléments

↓ corrections en opérant le regroupement d'éléments fils dans un enchâssant en demandant à chercher toutes les occurrences d'un élément au moyen du champ XPath et en tapant dans celui-ci « //Houssiaux » par exemple pour sélectionner chacun de ces éléments (éventuellement suivi d'un frère <Furne> et les insérer dans un élément parent <choix>,

↓ corrections encore en cherchant les éléments qui balisent des dénominations (Npersonne, etc.) au moyen du champ Xpath (« //Npersonne ») afin de contrôler chaque élément et de voir s'il faut déplacer ses balises pour inclure un choix,

↓ ajout de la déclaration de doctype avant la balise ouvrante de l'élément à la racine « <!DOCTYPE document SYSTEM "Sarrasine.dtd"> »

↓ reprise des dernières corrections jusqu'à conformité du XML avec votre DTD (*Sarrasine_***nom(s)***.dtd*),

*Sarrasine_***nom(s)***.xml*

.zip Les différents fichiers produits devront respecter les consignes de nommage et être réunis dans un dossier compressé de format ZIP déposé sur Moodle par l'étudiant du groupe dont le nom est le premier par ordre alphabétique :

📁 D1_ **nom(s)**.zip ← Sarrasine_ **nom(s)**.odt et Sarrasine_de_ODT_ **nom(s)**.xml
 ← Sarrasine_ **nom(s)**.xml et Sarrasine_ **nom(s)**.dtd

2. Balisage de *Sarrasine* en respect des recommandations de la *TEI:P5* [20 points]

[Travail en groupe de 1 à 3 étudiants (même groupe que pour la 1 partie ou en sous-groupes de 1 à 2 étudiants) pour le 11/01/2021]

Avec l'aide de la documentation en ligne de la *TEI:P5* (cf. *supra* et <http://www.tei-c.org/index.xml>), proposer un balisage respectueux des recommandations de la *TEI:P5* alternatif à notre balisage "personnel".

2.1. Production d'un XML conforme aux schémas de la *TEI* [12 points]

Pour cela, créer un document XML sur le modèle *TEI All* (menu Fichier > Nouveau > Cadre des modèles > TEI P5 > All [TEI P5]) donc ayant le contenu par défaut ci-après :

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml"
schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>édition Houssiaux (disponible en Wikisource) enrichie d'une comparaison textuelle avec l'édition Furne
(disponible dans le cadre du projet eBalzac</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

Compléter

– l'élément <teiHeader> en fournissant les métadonnées utiles au moins pour les éléments obligatoires

– et l'élément <text> avec le contenu de votre <document> initial où vous changerez vos noms d'éléments (et peut-être les enchâssements d'éléments) jusqu'à obtenir un document conforme au modèle de la TEI.

Si le travail de conversion du balisage personnel en balisage TEI est partagé entre plusieurs étudiants, il faudra que les conversions soient cohérentes entre elles et que les parties traitées par chacun aient les mêmes délimitations que précédemment (qu'elles soient clairement bornées au moyen de commentaires écrits sur les patrons suivants : « <!-- début du balisage par NOM_1 --> », puis « <!-- fin du balisage par NOM_1 et début du balisage par NOM_2 --> » et « <!-- fin du balisage par NOM_2 --> »).

2.2. Écriture d'une DTD aussi précise que possible pour ce document

Une fois le balisage respectueux des recommandations de la TEI terminé,

– supprimer les instructions de traitement qui imposent la conformité aux schémas de la TEI ainsi que l'attribut de l'élément <TEI> ;

– et ajouter la déclaration de DOCTYPE afin que le XML se réfère à une DTD *Sarrasine_comme_TEI_nom(s).dtd* que vous écrirez en essayant de proposer des déclarations d'éléments aussi précises que possible.

donc ayant le contenu par défaut ci-après :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI SYSTEM "Sarrasine_comme_TEI_nom(s).dtd"
<?xml-model href="http://www.tei-e.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.tei-e.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml"
schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-e.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>édition Houssiaux (disponible en Wikisource) enrichie d'une comparaison textuelle avec l'édition Furne
(disponible dans le cadre du projet eBalzac</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

Le XML TEI devra se conformer à votre DTD, mais sans modification du XML, c'est donc la DTD qui devra rendre compte de l'organisation du contenu du XML.

XML

M1 LCE ; M1 EL ; M1 LJ

7

SUJETS D'ÉVALUATION – © N. Gasiglia 04/12/2020

.zip Les différents fichiers produits devront respecter les consignes de nommage et être réunis dans un dossier compressé de format ZIP déposé sur Moodle (par l'étudiant du groupe dont le nom est le premier par ordre alphabétique) :

📁 **D2_***nom(s)*.zip ← *Sarrasine_TEI_**nom(s)*.xml [cf. devoir 2.1.]

← *Sarrasine_comme_TEI_**nom(s)*.xml

et

*Sarrasine_comme_TEI_**nom(s)*.dtd [cf. devoir 2.2.]